DOCUMENT RESUME

ED 385 546 TM 023 962

AUTHOR	Wilson, Kenneth M.
TITLE	Uses of the Secondary Level English Proficiency (SLEP) Test: A Survey of Current Practice.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-93-9; TOEFL-RR-43
PUB DATE	Mar 93
NOTE	86p.
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Educational Practices; Educational Research; *English (Second Language); Limited English Speaking; *Listening Comprehension; Listening Comprehension Tests; Reading Comprehension; Secondary Education; *Secondary School Students; Surveys; Testing; *Test Use; Test Validity
IDENTIFIERS	*Nonnative Speakers; *Secondary Level English Proficiency Test

ABSTRACT

The Secondary Level Proficiency (SLEP) test, offered by the SLEP School Services Program at the Educational Testing Service, measures English language listening comprehension and reading comprehension skills. It was developed for use with nonnative-English speaking students in grades 7 through 12. The SLEP is administered and scored locally, and the SLEP program does not receive routine feedback from local test users. This study was undertaken to obtain formal feedback from a sample of SLEP users through a survey questionnaire. Questionnaires were mailed to over 300 potential SLEP users. Although the return rate was relatively low (71 usable returns), the distribution of the returns by general institutional type and location was similar to that of the total sample. Survey findings provide information regarding: (1) testing practices; (2) purposes of testing; (3) selected characteristics of the examinees; (4) test-users' perceptions of the principal strengths and weaknesses of the SLEP and its manual; (5) the extent and nature of local studies concerned with validating the SLEP; and (6) related topics. Limitations of the findings for SLEP research are discussed. Four appendixes contain technical information about the survey and the questionnaire itself. Four exhibits, seven figures, and two tables illustrate the discussion. (Contains 18 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 385 546

TOEFL®

TEST OF ENGLISH AS A FOREIGN LANGUAGE

Research Reports

REPORT 43
March 1993

Uses of the Secondary Level English Proficiency (SLEP) Test: A Survey of Current Practice

Kenneth M. Wilson

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ETS
Educational Testing Service

M023902

BEST COPY AVAILABLE

USES OF THE SECONDARY LEVEL ENGLISH PROFICIENCY (SLEP®) TEST: A SURVEY OF CURRENT PRACTICE

Kenneth M. Wilson

Educational Testing Service
Princeton, New Jersey

RR-93-9

Educational Testing Service is an Equal Opportunity/Affirmative Action Employer.

Copyright © 1993 by Educational Testing Service. All right reserved.

No part of this report may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher. Violators will be prosecuted in accordance with both US and international copyright laws.

EDUCATIONAL TESTING SERVICE, ETS, the ETS logo, SLEP, TOEFL, and the TOEFL logo are registered trademarks of Educational Testing Service.

Abstract

The Secondary Level Proficiency (SLEP) test, offered by the SLEP School Services Program (SSP) at Educational Testing Service (ETS), measures English language listening comprehension and reading comprehension skills. It was developed for use with nonnative-English speaking students in grades 7-12. SLEP is administered and scored locally, and the SLEP® program does not receive routine feedback from local test users. The work described herein was undertaken to obtain formal feedback from a sample of SLEP users by means of a survey questionnaire. Questionnaires were mailed in April, 1991, to over 300 potential SLEP-use contexts worldwide (addresses of individuals, institutions or agencies placing orders for the SLEP or related materials within the most recent 18-month period). Although the return rate was relatively low (71 usable returns), the distribution of the returns by general institutional type and location was similar to that of the total sample. Survey findings provide information regarding testing practices, purposes of testing, selected characteristics of examinees (age/grade level, language background, and so on), test-users' perceptions of the principal strengths and limitations of the SLEP and/or the *SLEP Test Manual* (and suggestions for modification), the extent and nature of local studies concerned with validating the SLEP, and so on. Implications of the findings for SLEP-related research and development activities are discussed.

The Test of English as a Foreign Language (TOEFL®) was developed in 1963 by the National Council on the Testing of English as a Foreign Language, which was formed through the cooperative effort of more than thirty organizations, public and private, that were concerned with testing the English proficiency of nonnative speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program, and in 1973, a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the Graduate Record Examinations (GRE) Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; GRE Board members are associated with graduate education.

ETS administers the TOEFL program under the general direction of a Policy Council that was established by, and is affiliated with, the sponsoring organizations. Members of the Policy Council represent the College Board and the GRE Board and such institutions and agencies as graduate schools of business, junior and community colleges, nonprofit educational exchange agencies, and agencies of the United States government.

A continuing program of research related to the TOEFL test is carried out under the direction of the TOEFL Research Committee. Its six members include representatives of the Policy Council, the TOEFL Committee of Examiners, and distinguished English as a second language specialists from the academic community. Currently the Committee meets twice yearly to review and approve proposals for test-related research and to set guidelines for the entire scope of the TOEFL research program. Members of the Research Committee serve three-year terms at the invitation of the Policy Council; the chair of the committee serves on the Policy Council.

Because the studies are specific to the test and the testing program, most of the actual research is conducted by ETS staff rather than by outside researchers. However, many projects require the cooperation of other institutions, particularly those with programs in the teaching of English as a foreign or second language. Representatives of such programs who are interested in participating in or conducting TOEFL-related research are invited to contact the TOEFL program office. All TOEFL research projects must undergo appropriate ETS review to ascertain that the confidentiality of data will be protected.

Current (1991-92) members of the TOEFL Research Committee are:

James Dean Brown	University of Hawaii
Patricia Dunkel (Chair)	Pennsylvania State University
William Grabe	Northern Arizona University
Kyle Perkins	Southern Illinois University at Carbondale
Elizabeth C. Traugott	Stanford University
John Upshur	Concordia University

Acknowledgments

This report was made possible by the generous cooperation of individuals associated with programs of instruction in English for nonnative speakers of English, who responded to a survey questionnaire regarding the role of the Secondary Level English Proficiency (SLEP) Test in the programs involved.

Stella Cowell and her associates in the SLEP School Services Program, provided information that was essential in planning and implementing the survey, and helpful comments on various drafts of the questionnaire, as did Dorothy Bartram, TOEFL Agency Office, Canada. A number of other individuals took time from busy schedules to describe and discuss the programs for students of limited English proficiency with which they were associated (and in some instances also commented on a draft version of the questionnaire): Charles Huchet and his colleagues in Special Services, Princeton Regional Schools (NJ); Mark Tolo, English Program for International Students, Pennington School (NJ); Jay Doolan and colleagues, New Jersey Bilingual Programs; Donna Sue Pepperdine and colleagues, Yokohama Academy, Baltimore; Melissa Frimmer, Bloomfield Senior High School (NJ). Kimberly Graves, Temple University, Japan, and Rebecca Tillberg, Los Angeles Community College District, provided detailed information regarding SLEP use in their respective settings. Gerald DeMauro, Gordon Hale, and Donald Powers provided helpful reviews of drafts of this report. The foregoing contributions are gratefully acknowledged.

Table of Contents

Background 1
  Need for Feedback from Test Users. 2
  Purpose of the Present Study 2
  Questionnaire Development. 3
  Defining a Target "Population" for the Survey. 4
  Survey Mailing and Response. 5

Findings 6
  SLEP Use: Basic Data. 7
    Test Forms in Use. 7
    Extent of Reliance on SLEP 8
    Volume and Pattern of Testing. 8
  The Examinee Population. 9
    Age/Grade Levels of Examinees. 9
      Comments on "Out-of-Level" Use. 10
    Socio-Political Status of Examinees. 12
    Language Backgrounds of SLEP Examinees 14
  Purposes of Testing. 14
    Illustrative Elaborations of Reported Uses. 16
  Local SLEP Validation and Normative Studies. 18
    Local Studies and Local Norms. 19
    Evaluation of Respondent's Comments. 22
      Concurrent Validity 22
      Comments on SLEP's validity and usefulness . . . 23
      Establishing readiness for English-medium instruction 25
  Users' Perceptions of SLEP's Strengths and Limitations 27
  The "Ideal Test Package" 30
    More Comprehensive Assessment. 31
    Reduce Testing Time. 32
    A "Lower-Level" Test 32
    Freedom from Cultural and Gender Bias. 33

Summary. 34

Implications of the Survey Findings. 36
  Other Avenues for SLEP-Related R & D Activities. . . . 38
    Promote Widespread Use of Standard Procedures for Rating Productive Skills 38
    Explore SLEP's Validity Below the G7-12 Range. . . 38
    Increase SLEP's "Assessment Efficiency". 38
    Explore the Reliability and Validity of a Shorter Test 39
    Assess Contribution of Item Types to Validity. . . 39
    Enhance SLEP's General Diagnostic Potential. . . . 40
  Cultivate "Cooperative Interaction" with SLEP Users. . 40

Endnotes 65

References 71

List of Appendices

Appendix A: Illustrative SLEP Items 43

Appendix B: The Survey Questionnaire. 53

Appendix C: Procedures Followed in Modifying Existing Addresses for the Current Survey. 59

Appendix D: Illustrative Items from the Sequential Tests of Educational Progress (STEP): Listening Comprehension and Reading Comprehension 60

List of Exhibits

Exhibit A: Questions Regarding Local SLEP Studies and Norms Development 20

Exhibit B.1: Cover Letter, General Question. 54

Exhibit B.2: Questionnaire for U.S.A./Canada 55

Exhibit B.3: Questionnaire for Other Countries 57

List of Figures

Figure 1: Selected practices in using the SLEP test, without regard to type of use context . . . 8

Figure 2: Percentage of use-settings testing examinees at designated educational levels. 10

Figure 3: Reported status of SLEP examinees in the current survey: U.S.A./Canada (upper panel) vs. other countries (lower panel) 13

Figure 4: Principal language groups represented in SLEP use settings, and differences across settings in the linguistic diversity of SLEP examinees. 15

Figure 5: Purposes for which the SLEP test is being used, by location: U.S.A./Canada (solid bar) vs. other countries. 15

Figure 6: Have you conducted local studies? Developed local norms? Responses to Q5c and Q5e. . . 21

Figure 7: SLEP total scaled-score range of typical student at initial testing (Q5d). 21

List of Tables

Table 1: Types of Institutions/Agencies Identified as Having Placed Orders for the SLEP, By Location 5

Table 2: Distribution of Returns by Type of Use Context and Location. 7

Background

The Secondary Level English Proficiency (SLEP) test was developed by Educational Testing Service (ETS) to assess the English language listening comprehension and reading comprehension skills of " . . . students entering grades seven through twelve whose native language is other than English" (ETS, 1988: 5).[1*] More specifically, according to Stansfield (1984: 2) " . . . the test is designed for use as a selection or admissions instrument by private secondary schools, or as a placement instrument by public secondary schools." SLEP School Services Program publications (e.g., ETS, 1987) note that

> as a norm-referenced test, [the SLEP] provides users with the opportunity to compare student results with those of other students in similar situations. . .. A basic assumption underlying the SLEP test is that language ability is a critical factor in determining the degree to which secondary students can benefit from instruction; to succeed they must be able to understand what is being said (by both teachers and fellow students) and to understand both formal and informal material written in English (ETS, 1987: 5)

Users are informed (e.g., ETS, 1987: 5) that the SLEP® test can be helpful in making placement decisions such as, for example,

assignment to ESL classes,
placement in a mainstream English-medium program,
exemption from a bilingual program,
exit from an ESL program,
ESL program evaluation.

Although the SLEP test was initially developed for use with secondary-level (G7-12) student populations, based on information supplied by the program, the test is being used to assess the ESL listening and reading proficiency of nonnative-English speaking students at other age/grade levels (e.g., 6th grade students, college-level ESL students), academically unclassified adults (e.g., enrollees in English-language institutes, adult ESL classes); and so on.

Three statistically equivalent forms of the SLEP test are offered through the SLEP School Services Program: Form 1, developed in 1979-80; Form 2, developed in 1980-81; and Form 3, developed in 1986-87. Each form is made up of 150 multiple-choice questions of eight different types (see

---

* See numbered endnotes.

Appendix A). Test booklets are reusable; examinees use separate answer sheets to record their responses to the test questions.[2]

Various editions of the *SLEP Test Manual* (see, for example, ETS, 1988; 1991) provide, among other things, information regarding the psychometric characteristics of the test, evidence of validity (e.g., systematic differentiation of groups classified according to ESL proficiency levels, relatively high correlation with the TOEFL, and so on), general guidelines for test use and interpretation, and suggestions for local research (e.g., it is recommended that users conduct studies designed to assess the extent to which SLEP scores are related to self-assessments, teachers' ratings of ESL proficiency, or performance in regular academic courses).

Need for Feedback from Test Users

Because the SLEP test is locally administered and scored, the SLEP School Services Program does not routinely receive information pertaining to test use, local studies, and other related matters. The program also does not receive the kind of examinee-generated feedback that is routinely available for a centrally administered test such as the TOEFL--for example, examinees' test scores, item-level responses, answers to background questions.

Without such information, the program is limited in its ability to judge the extent to which current forms of the SLEP test are meeting the ESL assessment needs of practitioners in diverse settings, to introduce modifications that may be needed to improve the overall usefulness of the test, or to routinely summarize, evaluate, and publish data on test performance for various subgroups (e.g., age/grade level, language background).

Purpose of the Present Study

The work described herein was undertaken to obtain formal feedback from users of the SLEP test by means of a survey questionnaire concerned with matters such as those alluded to above. More specifically, the survey was designed to obtain information bearing on the following general lines of inquiry:

- What are the basic patterns of test use (e.g., test forms used, number of examinees tested, number of times each examinee is tested, other assessment procedures used in conjunction with the SLEP)?

- What are the characteristics of examinee populations (e.g., age/grade levels, socio-political status [e.g., refugee, immigrant, international student], language background)? If the test is being used with examinees not classifiable within the G7-12 range (e.g., 6th graders, college-level students, older adults), what are the judgments of test users regarding the test's suitability or lack of suitability for such use?

- To what extent is the SLEP being used for purposes suggested in the *SLEP Test Manual* (e.g., assessment of the readiness to undertake English-medium academic instruction, placement for ESL/EFL instructional purposes, program evaluation, admission, monitoring the progress of individual students, and so on).

- Are test users conducting local studies of the relationship between SLEP scores and direct measures of ESL/EFL students' ability to use English (e.g., teachers' ratings)? Are they developing local norms, as suggested by the program? What is the scaled-score range that includes the average total score obtained by students when initially tested?

- What are the principal strengths and limitations of the SLEP Test and/or the *SLEP Test Manual*, from the perspective of local test users? What changes or modifications, if any, do users recommend?

- Generally speaking, what characteristics of a standardized test of ESL/EFL proficiency (and related developer-produced materials and services) do test users believe would be most helpful/useful in use contexts similar to their own?

Questionnaire Development

The foregoing questions were judged to be generally applicable for test users regardless of location (that is, whether inside or outside the United States) and type of setting (e.g., school, college, language institute).[3]

A draft questionnaire that included both precoded and open-ended response options was developed, in consultation with program staff, and pretested.[4] Based on results of pretesting, it was decided that the one basic set of questionnaire items would be appropriate for all test-use contexts, with only minor changes in wording--primarily in connection with certain testing procedures that are mandated by statute in the U.S. and Canada, but not elsewhere.[5]

The two versions are shown in Appendix B, along with the (undifferentiated) cover letter that accompanied each questionnaire (cf. options for Q1 and Q6 in the respective versions of the questionnaire).

Defining a Target "Population" for the Survey

Orders for the SLEP test are received from diverse institutions and agencies, as well as from professionally qualified individuals, in the United States and elsewhere in the world. These include public schools, private academies or preparatory schools, international student exchange programs, language institutes, corporations; postsecondary institutions located in the U.S., Canada, and elsewhere.

The possibility of identifying, evaluating, and surveying the total population of institutions, agencies, or individuals ordering (using) the SLEP test during the past decade was considered. However, it was not feasible to undertake the substantial effort that would be needed to retrieve and match order-files across successive fiscal cycles during the decade.[6]

After evaluating and ultimately rejecting this comprehensive approach, it was decided to survey a sample definable on the basis of records included in the "current" systems file--typically covering transactions over approximately the most recent 18 months--for the period, roughly, between July 1989 and December 1990.

Computerized printouts of addresses representing distinct transactions (that is, one or more orders for SLEP-related materials) during the period were used to identify the ordering institutions/agencies/individuals; additional addresses were supplied by the TOEFL representative office in Canada.[7] This process resulted in the identification of 356 different "potential" SLEP-use contexts (that is, different purchasers of the SLEP test and/or related materials).

Based on the basic identification provided in the fiscal files, these potential-use contexts were classified as being primarily, (a) academic (secondary vs. postsecondary), (b) language institutes, (c) international student exchange programs, or (d) corporations or business institutions. The distribution of these potential-use settings by type and location is shown in Table 1.

It can be seen that most of the orders (about 84 percent) were shipped to academic settings--some 59 percent classified as "secondary-level" and 25 percent classified as "postsecondary-level." Approximately 71 percent of the orders were

TABLE 1. Types of Institutions/Agencies Identified as Having Placed Orders for the SLEP, By Location

Type*	Location: U.S.A	Canada	Other	Total	Percent
Academic, Total	201	49	48	298	(83.7)
Secondary	147	25	27	209	(58.7)
Postsecondary	54	14	21	89	(25.0)
Exchange Program	11			11	(3.1)
Language Institute	5	1		6	(1.9)
Corporation	6			6	(1.9)
Other**	20	5		25	(9.8)
Total (All types)	253	55	48	356	
Percent	71.1	15.4	13.5		(100.0)

* Classification inferred from information contained in the addresses to which shipments of SLEP test-booklets and related materials were mailed.

** This category includes orders from institutions or agencies not clearly classfiable according to the preceding categories (e.g., unfamiliar academic identification, embassies, correctional institutions, individuals without institutional identification, and so on.

shipped to addresses (including APO/FPO) in the U.S., 15 percent were sent to Canadian addresses, and the remaining 14 percent to all other addresses.

About 10 percent of the sample could not be classified with certainty according to one of the specific categories indicated (e.g., governmental agencies; individuals with professional, but not institutional, identification; unfamiliar acronymic designations; and so on).

Survey Mailing and Response

Survey questionnaires were mailed on April 19, 1991. No final reply date was specified. During the first four weeks, returns were limited in volume and scattered (that is, there was no clearly discernible peak). Both the timing of the survey (coinciding with end-of-school-year pressures) and the lack of consistent personal identification (e.g., name, title, and program) for the person actually responsible for test use, militated against the prospect of substantially increasing the

overall response through followup mailings, and none were made (see Appendix C for procedures used in an effort to "personalize" the basic mailing). However, returns were not formally "closed" until September 30, 1991.

As of that date, a total of 71 completed questionnaires had been received, distributed by type-of-use context and location as indicated in Table 2; approximately 30 percent were received after June 30, 1991. The marginal distributions of returns by type- and location-of-use context, shown in Table 2, were similar to the distributions that were obtained for the total survey sample (Table 1). It thus appears that the responding sample is reasonably representative of the total sample with respect to both type and location of test-use contexts.

In three instances, two completed questionnaires were returned in the same envelope: one set from two ESL teachers in different high schools in the same school district, one set from two admissions office staff members in a preparatory school, and a third set from two members of the ESL program staff at a university in the United States. Both questionnaires in each set were processed without special treatment.

In addition to the completed questionnaires, five questionnaires were returned unopened (for insufficient address), and five were returned not fully completed. Responses to precoded items were keyentered. Verbatim copies of write-in responses were prepared to facilitate evaluation of comments, suggestions, and recommendations from respondents. Moreover, respondents who provided information suggesting that systematic local studies of the concurrent or predictive validity of SLEP test scores had been conducted, were contacted (by letter, FAX, and/or telephone) in an effort to obtain additional detail.

Findings

Survey findings, summarized below, provide information regarding (a) the scope, volume, and frequency of testing with the SLEP, (b) characteristics of examinee populations in various use contexts, (c) the purposes for which the SLEP is being used, (d) the extent to which SLEP users are conducting local validation studies and/or developing local norms, and (e) respondents' perceptions of the most positive aspects and the principal limitations of the SLEP and/or the *Test Manual*; their suggestions for change; and their characterizations of the hypothetical ESL proficiency test (and test-developer provided services) that would be most useful in contexts such as their own.

TABLE 2. Distribution of Returns by Type of Use Context and Location

Type of use context	Location U.S.	Canada	Other	Total	Percent	(Mail)*
Academic	41	4	11	58	81.7	(84%)
Secondary	35	3	7	45	63.4	(59%)
Postsecondary	6	3	4	13	18.3	(25%)
Language Institute**	3	1		4	3.9	(4%)
Corporation**	2			2	1.7	(2%)
Exchange**	5		2	7	10.0	(4%)
Total	51	7	13	71		
Percent	72%	10%	18%	100%		
(Mail)	71%	15%	14%	100%		

* Entries in this column do not total 100% due to the fact that several returns were received from representatives of institutions/agencies not classifiable as academic, exchange, language institute, or corporation on the basis of information available when questionnaires were mailed (that is, returns from "Other" in Table 1).

** Type of use context was reported by respondents (see cover page of questionnaire), but academic level was inferred from other information available.

Some implications of the findings are considered in the final section.

SLEP Use: Basic Data

Figure 1 shows percentage distributions of responses to questions about (a) forms of the SLEP currently in use, (b) extent of reliance on SLEP only vs. SLEP in combination with other ESL assessment procedures, (c) the number of individual examinees tested and the typical number of times each examinee was tested during the most recent 12 month period.

Test Forms in Use

Forms 1 and 2, but not Form 3, were reported being used by a majority of the respondents. Almost 90 percent reported using Form 1, 70 percent reported Form 2, and 37 percent reported Form 3.

BEST COPY AVAILABLE

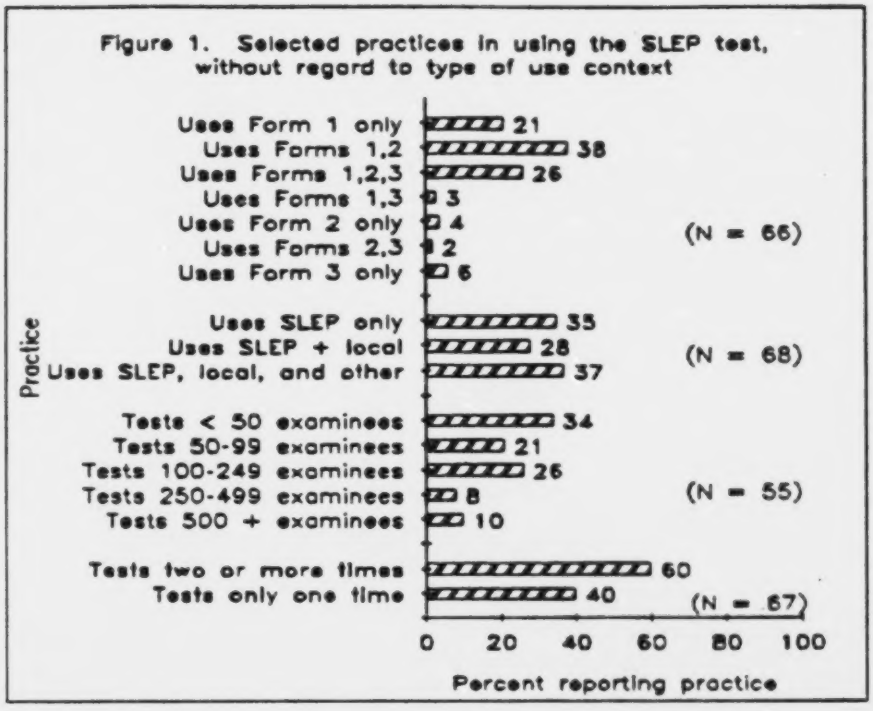


Figure 1. Selected practices in using the SLEP test, without regard to type of use context

Extent of Reliance on SLEP

One-third of the respondents reported that they used only the SLEP for ESL assessment; about two-thirds reported that the SLEP test was used in conjunction either with local assessment procedures only (28 percent) or with local procedures, plus one or more additional standardized ESL tests (37 percent). The number of individuals who responded to each of the questions involved is shown as the base for percentages.

Volume and Pattern of Testing

The number of examinees tested with the SLEP varied markedly across use contexts, ranging from less than 10 annually to 1,000 or more.[8] However, about 80 percent of those who supplied pertinent information (only 55 of 71 did so) reported testing fewer than 250 individuals, and a majority tested fewer than 100. Some 60 percent reported that the typical individual was tested two times, and 40 percent reported only one-time testing.

The Examinee Population

The SLEP test was originally developed for use with a population of examinees made up primarily of nonnative-speaking international students who need to demonstrate their ESL proficiency in connection with plans to enter an English-medium secondary school (G7-12) program in the United States or elsewhere. However, the test has been used not only with G7-12 students, but also with 6th graders and postsecondary-level students.[9]

The test-taking population also appears to include some nonnative-speakers who are not in "regular academic progression" as international students planning to study in an English-medium environment--e.g., political or economic refugees, immigrants, and so on. Members of these groups may differ from "regular students" with respect to age, educational level, English-language background, and other variables.

Accordingly the survey contained questions designed to assess

(a) the extent to which SLEP is being used at various age/grade levels (percentage of examinees who are below 7th grade, in the G7-12 range and beyond the G7-12 range, respectively);

(b) users' assessments of suitability/unsuitability for examinees below or above the G7-12 age/grade range; and

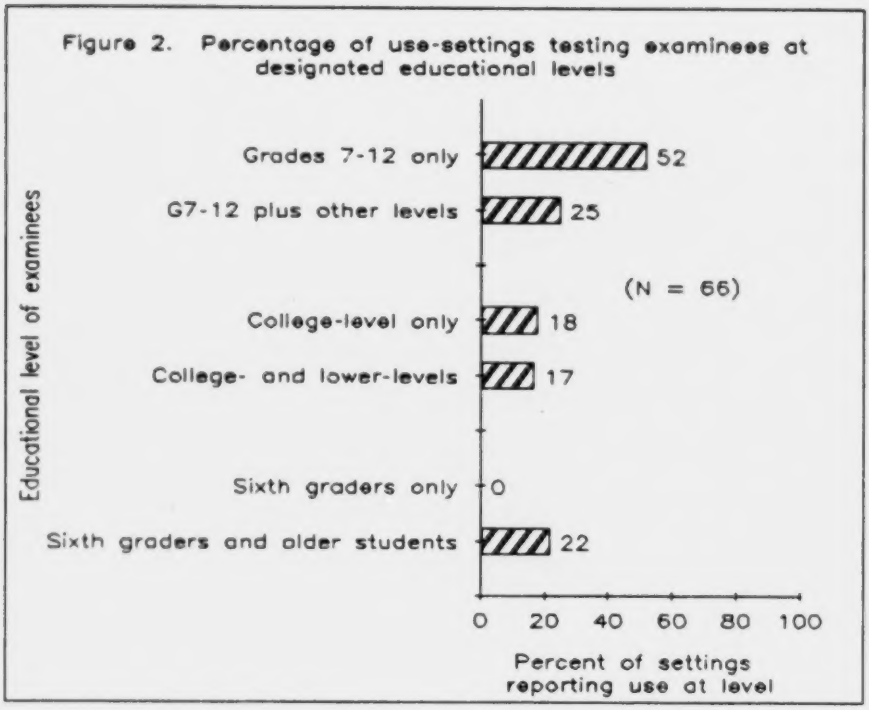
(c) the socio-political status of the students involved (percentage of examinees in designated categories).

Age/Grade Levels of Examinees

As may be seen in Figure 2, more than three-fourths of the respondents were testing at least some students in the G7-12 range for which the SLEP was originally designed. However, about 35 percent were testing <u>some</u> postsecondary level students, and about one-fourth were testing <u>some</u> students below the 7th grade level.

- Testing was restricted to students in the G7-12 range <u>only</u>, in only 52 percent of the settings that reported this information; some 18 percent reported testing <u>only</u> postsecondary level students.

Respondents indicating that the SLEP was being used with examinees whose age/grade placement was either lower or higher than the originally targeted G7-12 range were invited to



comment on the test's suitability for the younger or older examinees involved, and to provide specific examples. A total of 23 respondents commented.

The comments, by and large, were general appraisals that did not specify particular test characteristics or provide specific examples. Only the "map" items (see Appendix A) were singled out by several of the respondents as being inappropriate or too difficult for 6th graders and other examinees.[10]

The general flavor of the comments is captured in the verbatim excerpts that follow (emphasis added in all instances).

Comments on "out-of-level" use. An American International School, South America (10 percent 6th graders).

> *New students whose first language is not English are tested for placement (ESL I [beginners]/ESL II [intermediate-advanced] or regular class). Students placed in ESL are retested in the middle and at the end of the year to assess*

progress and decide when they should be mainstreamed into the regular classes. We used it for 6th graders as well. It appears to be suitable.

Oregon, Public Middle School (20 percent 6th graders).

We administer the SLEP test to ESL students once a year. From the results I can decide whether or not the student should have to take the standardized achievement tests and also what support services are needed in academic areas. The sixth graders had no problems. They were on-task and tried to do their best. The older students (adult ESL) had never filled in an answer sheet of any kind, so I gave extra instruction and prompting as needed. All the students fail miserably on the map section.

Virginia Public School (occasional 6th grader).

Content better suited to 7-12 graders than to sixth graders. Map test is definitely geared to older students. However, a great test.

Oregon, Educational Service District (some 6th graders).

We teach ESL in migrant/bilingual resource rooms in two counties in rural eastern Oregon. The SLEP is used each spring to evaluate growth in our secondary students (G7-12). ESL is an elective class where students are enrolled in one or two periods per day. 7th and 8th grades do well; 6th graders do not do well; listening part of test is good.

Michigan, Middle School (G6-8).

Test is administered in the fall (form 1) and spring (form 2) to evaluate growth and need to be in ESL class. Listening comprehension: Map and cars unsuitable. This section seems to be particularly confusing to the students because they are not familiar with the concept of driving.

Louisiana, University-Based Intensive ESL Program.

SLEP used to place students in our month-long intensive program. Students are tested when they first arrive in the program and are placed solely on their SLEP score. It is not used for post-testing or advancement. Most of our students are 18-26 and some (approximately 20 percent) 26 and older. Age does not appear to be a factor, except with much older students who appear to be intimidated by standardized tests.

Japan, American Liberal Arts College.

> *We are an American liberal arts college operating in Japan. We use the SLEP as one factor in determining admissions, placement, and promotion. Most of our students are between the ages of 18-21. These students have completed high school, and many have attended special schools for one additional year trying to get into university. SLEP is helpful in determining general language abilities.*

Japan, U.S. University Branch.

> *Virtually all of the students tested with the SLEP exam are in the 18-20 year age range. The exam is fairly well-suited to our student applicants, but is perhaps a little more suitable for a slightly younger age group.*

U.S. University.

> *There are problems with content. . .; the car map is confusing in test form 1.*

Japan, U.S. University Branch.

> *We use the SLEP as part of the admissions assessment for Japanese students entering our IEP. The main purpose is for placement of students in roughly equivalent groups. The SLEP gives a good general assessment of student achievement levels.*

On balance, the comments suggest that the SLEP is perceived to be "suitable" for use with examinees at quite diverse educational levels from the 6th grade through, at least, the early postsecondary years. It is also perceived as being, in some ways, possibly less suitable for younger students (below the 7th grade level) than for postsecondary level students.

Socio-Political Status of Examinees

Figure 3 shows distributions of means of reported percentages of examinees in designated sociopolitical categories, as reported by survey respondents in the U.S., Canada, and other countries. The several distributions shown in the figure are generally similar to the comparable distribution reported by Stansfield (1984) for the basic SLEP reference group--a sample of ESL students in U.S. secondary schools.

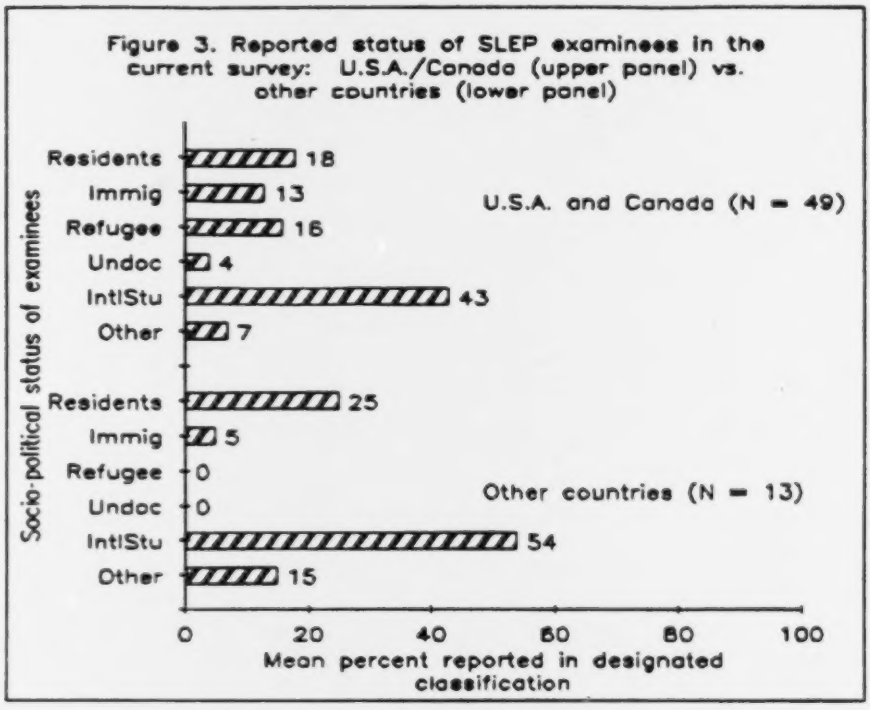


Figure 3. Reported status of SLEP examinees in the current survey: U.S.A./Canada (upper panel) vs. other countries (lower panel)

In the reference group sample, approximately one-fourth of the examinees involved were self-classified as foreign students; slightly lower percentages were classified as immigrants, refugees, and U.S. citizens, respectively.

In the current survey sample, respondents from the U.S.A./ Canada reported an average of slightly more than 40 percent international students, compared to an average of 50 percent in this category reported by respondents in other locations. Examinee populations that included refugees or undocumented individuals were largely restricted to settings in the United States and Canada.

Some indication of the types of demographic diversity represented in SLEP-use settings is provided by the following descriptions.

BEST COPY AVAILABLE

Canadian College.

> *The age range is 18 to 70, though most students are in their 20's;. . . classes for beginners with virtually no English, right up to college prep; some are refugees with very little formal schooling; others have the equivalent of high school in their mother tongue, many have university backgrounds in their mother tongue, but function in English at very low level.*

California School District.

> *About 1,000 (G7-12) students tested last year; 70 percent Spanish, 50 percent undocumented, 30 percent refugees, 20 percent recent immigrants.*

Language Backgrounds of SLEP Examinees

Most of the respondents (67 of 71) answered a question regarding the language backgrounds of examinees tested with the SLEP. Eleven (11) language groups were designated on the questionnaire; respondents were asked to write in names of other pertinent groups. Figure 4 shows the percentage of respondents reporting each of the 11 designated groups and the distribution of use contexts according to the total number of different language groups reported (designated plus write-in).

The data in Figure 4 simply point up the language groups that are most consistently represented across SLEP-use settings, and indicate that the examinee populations served by SLEP-use contexts differ considerably with respect to degree of linguistic heterogeneity. In about one-third of the SLEP-use settings, for example, only one language group is being tested, whereas at the other extreme, one respondent reported more than 50 "nationality groups."

Purposes of Testing

As may be seen in Figure 5, test users in the United States and Canada (solid bars in the figure) and their counterparts elsewhere in the world (hatched bars), reported a generally similar pattern of testing purposes. Only a few respondents (about 11 percent) reported testing for only one of the purposes designated in the questionnaire.[11]

- Assessing the readiness of ESL students for English-medium academic instruction was the most frequently reported purpose for testing. This purpose was cited by more than two-thirds of all respondents (U.S.A./Canada, 64 percent; other countries, 77 percent).

Figure 4. Principal language groups represented in SLEP use settings, and differences across settings in the linguistic diversity of SLEP examinees

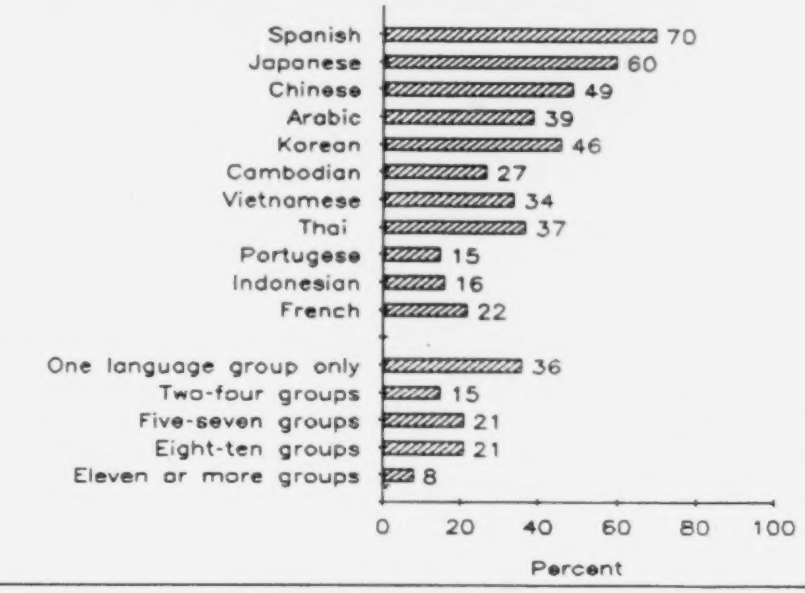
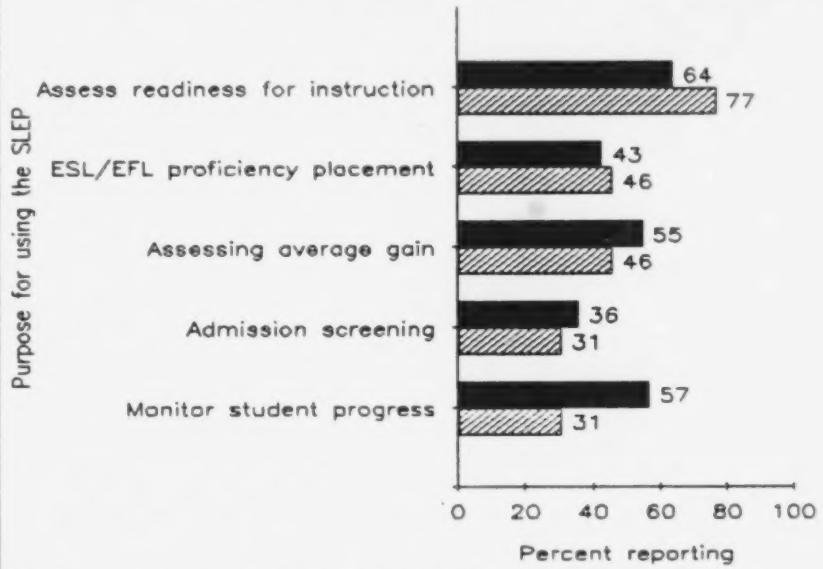


Figure 5. Purposes for which the SLEP test is being used, by location: U.S.A./Canada (solid bar) vs. other countries


Purpose for using the SLEP
Assess readiness for instruction 64 77
ESL/EFL proficiency placement 43 46
Assessing average gain 55 46
Admission screening 36 31
Monitor student progress 57 31
0 20 40 60 80 100
Percent reporting

Note. U.S.A./Canada (N = 56); Other (N = 13); NR = 2.

BEST COPY AVAILABLE

- Use of the SLEP in assessing average (net) gain in proficiency after instruction in EFL, ESL, Bilingual, and similar programs, using either locally devised models or mandated evaluation models (U.S.A/Canada), was reported by a slight majority of respondents overall.

Data not shown in the figure indicate that about 33 percent of the U.S.A./Canada respondents in secondary-school settings reported using the SLEP to assess " . . . average gain (using mandated evaluation models) in programs defined/required by statute (Federal/State/Provincial)." Gain assessment, without regard to model, was cited by about 75 percent of U.S.A./Canada respondents in secondary school settings.

- Placement of students for purposes of ESL or EFL instruction was reported to be a testing objective by more than 40 percent of the respondents.

- Screening for admission to an institution or program was cited as a purpose for testing by about one-third of all respondents. Data not shown in the figure indicate that this purpose was cited by all the exchange-program respondents.

- Monitoring the progress of individual students was cited by some 57 percent of U.S.A./Canada respondents and about 31 percent of other respondents.

All of the foregoing, of course, are well-established objectives of ESL proficiency testing; several illustrative descriptions provide a more detailed perspective.

Illustrative Elaborations of Reported Uses

U.S. Private Secondary School.

International students are given the SLEP test and another assessment tool to determine whether they are proficient enough in English to be placed in the regular academic curriculum or the ESL program. Based on these test scores, students are then divided into four levels of proficiency and are placed in courses according to these levels in the ESL program. Students are tested again in December and in May to determine their progress in English.

U.S. Public Secondary School for International Students.

Used for entrance screening to assess levels of English proficiency to determine whether we will accept them into the academic program. Also for measuring progress on a yearly basis, and deciding which ESL classes they should take.

Intensive ESL Program (U.S. University).

We use the SLEP to place students in our month-long intensive program. Students are tested when they first arrive in the program and are placed solely on their SLEP score. It is not used for post-testing or advancement. Some 65 percent of students are short-term language students planning to return to their country upon completion of studies at the language institute.

Exchange Program (U.S.).

(We) use the SLEP test to assess the level of English proficiency of our high school aged foreign exchange students. All foreign students are given the SLEP test prior to their acceptance to participate in an academic year in the United States. Our overall objective is to assess each student's English ability and their ability to function in an English-speaking high school. We use the SLEP test to screen our students for acceptance to the program (who must achieve a minimum score for acceptance).

U.S. Public School District.

We teach ESL in migrant/bilingual resource rooms in two counties in rural eastern Oregon. The SLEP is used each spring to evaluate growth in our secondary students (grades 7-12). ESL is an elective class where students are enrolled for one or two periods per day. The class is graded and carries high school credit (grades 9-12). As required by our Migrant and bilingual Federal Program, we test every spring all secondary students being served in an ESL component. Teachers do informal assessments for their own diagnostic purposes. We've reported the scores to our program evaluator. The teachers use the results for their information informally only.

U.S. College (Japan).

We use the SLEP results to help decide on placement level for applicants wishing to enter our intensive English Language Program, ... In addition to the SLEP, we also administer a 25-minute English essay exam and conduct a 10-minute oral interview (with a trained ESL professional) for each applicant. The SLEP results comprise one-half of the overall result, while the essay score and interview score each comprise one-fourth.

BEST COPY AVAILABLE

U.S. Independent Secondary School.

We use the SLEP for all our new students from foreign-speaking backgrounds for placement in three different English classes. Then we give them the test again at the end of the year to assess the gains they have made. We also use it for admissions guidance if the applicant has not taken the TOEFL. The SLEP is studied after administration and items are studied to ascertain weaknesses that can then be worked on in class.

U.S. Middle School.

I administer the SLEP test to ESL students once a year. From the results I can decide whether or not the student should have to take the standardized achievement tests and also what support services are needed in academic areas. SLEP is a standardized report card to show other school personnel that regular testing is appropriate or inappropriate with individual students.

U.S. School District.

The . . . Unified School District uses only the reading portion of the SLEP test for students enrolled in the Secondary Level ESL Program at middle schools and high schools. The assessment is given to determine entry and exit level reading skills. The reading portion of the SLEP test is used as 1) an initial or entry assessment to determine a student's reading proficiency and 2) an end-of-the-year assessment to determine the student's progress.

U.S. Public School.

SLEP is used to assess listening and reading comprehension of ESOL students in 7-12, fall and spring of each year. The scores are used (along with other testing data) to determine placement and exit of ESOL students in the ESOL program. (Especially reading suitable for LEPs).

Local SLEP Validation and Normative Studies

Through the *SLEP Test Manual* (e.g., ETS, 1991, 1988, 1987), the School Services Program advises test users (a) to conduct local studies designed to assess the relationship of SLEP scores to teachers' observations of proficiency and other pertinent performance criteria, and (b) to develop local norms.

SLEP users were asked to indicate whether they had conducted studies along lines indicated above and developed local norms, respectively. They were invited to provide brief descriptions of any studies that may have been conducted, or,

in the absence of formal studies, to indicate their impressions of the relationship between SLEP scores and direct measures of proficiency. They were also asked to supply norms tables, if available.

The specific questions posed for consideration by respondents are shown in Exhibit A. Selected SLEP reference-group "placement" information has been inserted opposite the respective SLEP total score ranges (Q5d in Exhibit A). This information was not included in the survey questionnaire itself.

It was assumed that most respondents--including those who may not have conducted formal studies or developed local norms--would be able to indicate the score-range that included the average total scaled-score for examinees taking the test locally. It was hoped that some users would be in a position to forward reports of local studies that would provide descriptive statistics for defined subgroups (age/grade level, years of ESL/EFL study, and so on) and other validity-related evidence. However, they were not directly invited to do so.

Local Studies and Local Norms

It can be seen in Figure 6 that, of 71 test users responding to the survey, about 60 percent reported conducting local studies of the type described in Q5c, but only 14 percent indicated that local norms had been developed.

Taken at face value, these figures suggest widespread lack of attention to the development of local norms, as defined in the questionnaire, namely, as "a table showing the percentage of students scoring at or below designated SLEP scores." Only 14 percent of the respondents reported having developed such tables. None of the respondents supplied a norms table meeting the definition involved, although specifically invited to do so.[12] By inference, locally developed tables of this type are not "essential" for local testing purposes in contexts such as those represented in the survey sample.

Notwithstanding apparent lack of attention to "local norms development," a substantial majority of respondents provided information regarding the average SLEP performance of their students at the time of initial testing (Question Q5d), as can be seen in Figure 7. Even so, 22 percent either did not respond at all or indicated two or more score-categories (included in the NR category).

It is apparent from the distribution of reported averages in Figure 7, that the SLEP is being used with local examinee-populations that differ markedly, on the average, with respect to level of developed ESL proficiency.

Exhibit A

Questions Regarding Local SLEP Studies and Norms Development

Q5c. Have you been able to study the relationship between SLEP scores and direct measures of ESL/EFL proficiency (for example, ESL/EFL instructor's ratings of oral English proficiency; academic teachers' ratings of students' ESL/EFL skills)?

____1. Yes (please describe briefly)

____2. No (please comment briefly on your impressions regarding the foregoing, and reasons for them).

Q5d. Five SLEP total scaled-score categories are specified below. Please check the score-range that includes the average score typically obtained by students when initially tested.

(SLEP reference group placement)*

1. < 33 = P24 (Bilingual, Full-time, Mean = 32)
2. 33-39 = P39 (Bilingual P-T or ESL F-T, Mean = 37)
3. 40-46 = P57 (ESL Part-time, Mean = 43)
4. 47-53 = P75 (Mainstream Class, Mean = 50)
5. 54 + = > P75 (No subgroup at this level)

Q5e. Have you developed local norms for the SLEP (e.g., a table showing the percentage of students scoring at or below designated SLEP scores)?

____1. Yes (if possible, please enclose a copy of your norms table and related description)

____2. No

---

* The "SLEP reference group placement" data (percentile ranks for upper-limit of score intervals, and total score means for placement levels) included with Q5d above, reflect findings of the initial SLEP validation study (Stansfield, 1984; also reported in various editions of the *SLEP Test Manual* [e.g., ETS, 1987]). These data were not included as part of the basic question posed for survey respondents.

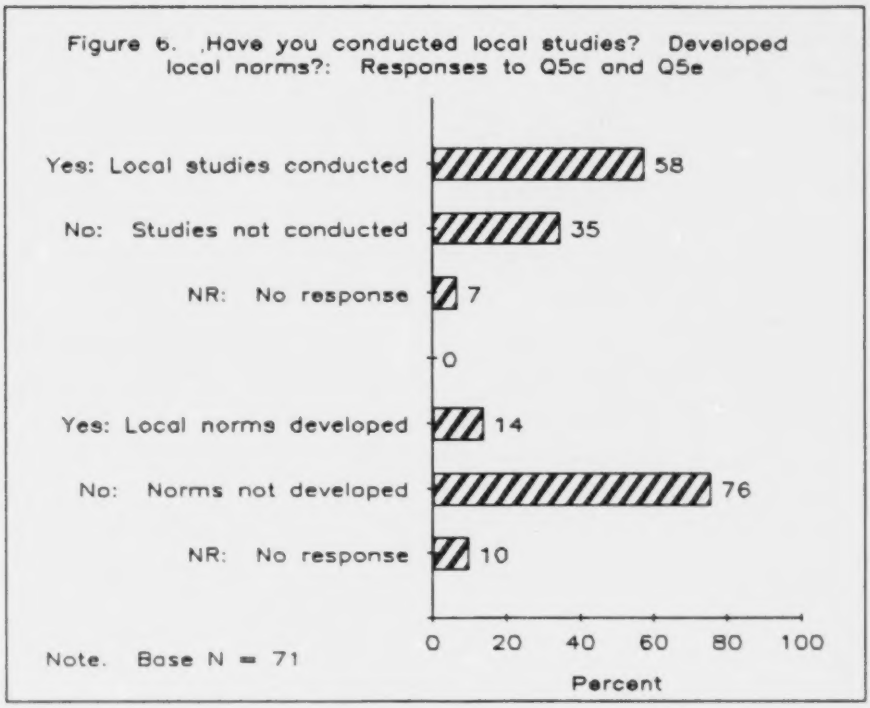


Figure 6. Have you conducted local studies? Developed local norms?: Responses to Q5c and Q5e
Yes: Local studies conducted
58
No: Studies not conducted
35
NR: No response
7
0
Yes: Local norms developed
14
No: Norms not developed
76
NR: No response
10
Note. Base N = 71
0 20 40 60 80 100
Percent

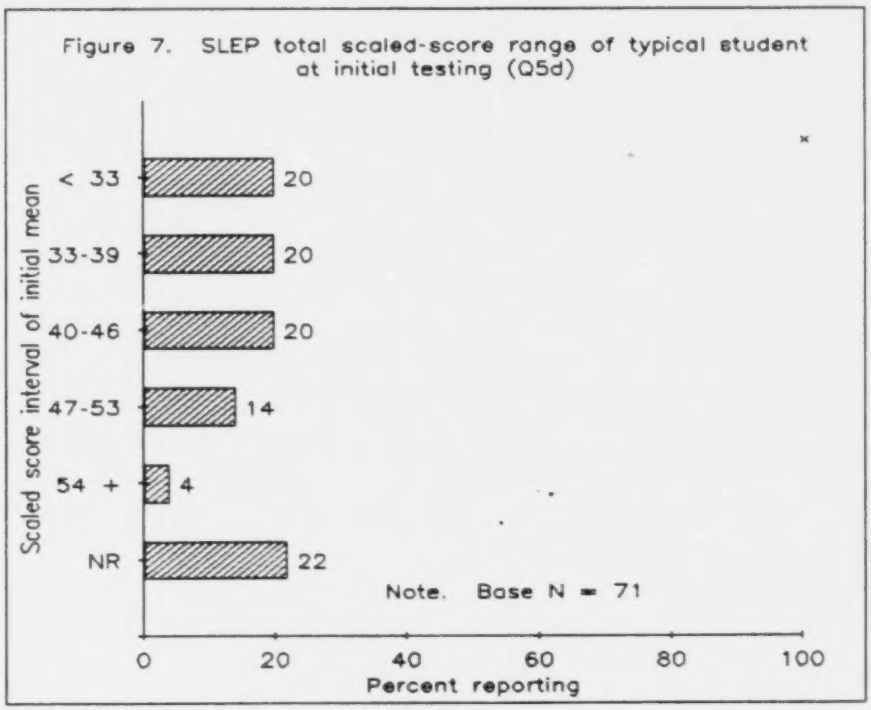


Figure 7. SLEP total scaled-score range of typical student at initial testing (Q5d)
Scaled score interval of initial mean
< 33
20
33-39
20
40-46
20
47-53
14
54 +
4
NR
22
Note. Base N = 71
0 20 40 60 80 100
Percent reporting

Some sense of the functional implications of differences in average SLEP performance is conveyed by the proficiency-placement levels designated in Exhibit A, above, with the corresponding mean total SLEP scaled scores for students in the respective levels (from Stansfield, 1984; also reported in various editions of the *SLEP Test Manual* [for example, ETS 1987, 1988]).

Evaluation of Respondents' Comments

A majority of the respondents accepted the invitation to "please describe briefly" any local studies that may have been conducted or to indicate their impressions of the relationship between SLEP scores and direct measures of proficiency as outlined in Question Q5c.[13]

Both the descriptions of local studies and study findings, and informal observations regarding SLEP's concurrent or predictive validity, varied markedly in style as well as substance--by inference, reflecting similarly marked differences in the nature, scope, and degree of psychometric and statistical rigor of the local studies involved.

Almost one-half (33 of 71, or 46 percent) of survey respondents (24 who did not report a study and 9 who did so) neither commented on SLEP's concurrent validity nor provided information bearing directly or indirectly on SLEP's validity for local purposes.[14]

The remaining survey respondents (38 of 71) provided comments, not all of which were deemed to be directly responsive to the question posed.[15] Most of the responsive comments involved direct or indirect allusions to SLEP's relationship to other measures, or SLEP's usefulness or lack of usefulness for local purposes--e.g., placement, including references to score levels at which students are judged to be ready to enter full-time English-medium instruction.

Concurrent validity. The most comprehensive program of local validation research described by a respondent to the survey, involved the systematic assessment of concurrent relationships between SLEP scores and direct assessments of oral English proficiency and writing skills, respectively, in samples of Japanese students in the intensive English Language Program of the Japanese branch of a U.S. university.

Testing Director, U.S. University (Japan).

> *We use the SLEP results to help decide on placement level for (such students). In addition to the SLEP, we also administer a 25-minute English essay exam and conduct a 10-minute oral interview (with a trained ESL*

professional) for each applicant. . . . Our exam is locally developed and holistically scored by two readers on a 1-6 scale; our interview 'test' is also locally developed--students are rated on a scale of 1-7 in six areas of communicative behavior.

Over a period of two academic years, from November '89 to April '91, the three forms of the SLEP were administered a total of 12 times as part of our IELP placement tests. I have observed a Pearson correlation coefficient ranging from .57 to .72 and averaging .63 (N = 1,648), with writing exam scores . . ., and a correlation coefficient range of .55 to .69, averaging .63, with interview scores (in samples with initial total scaled score means in the 40-46 range).

The studies just described reflect an unusually thorough and comprehensive application of indirect and direct measures in placing students--use of a composite score derived from the SLEP, the oral proficiency rating, and the essay rating.

The information supplied by other respondents who commented on SLEP's validity and usefulness was not buttressed by citation of empirical findings comparable to the foregoing. At the same time, there was a relatively consistent "positive validity" theme in the comments--that is, relatively consistent reports, based on formal and informal observation, of positive relationships between SLEP scores and more direct measures such as those referred to in the question, and/or statements indicating that the SLEP had been found to be "valid" or "useful" for local purposes.

Verbatim excerpts from all the comments that were deemed responsive to the request for information about studies of SLEP's validity (see note 12, above, and related discussion), by respondents who reported that a study had been conducted, reflect the general themes outlined above. A few individuals offered comments bearing on SLEP's validity or usefulness, based on informal observation only. These comments are identified accordingly in the summary statements that follow.

Comments on SLEP's validity and usefulness. Oregon Public School.

Teachers report that rankings of students by the SLEP generally reflect their own assessments. The correlation between reading scores on the SLEP and district graduation standards is .58.

English Language Institute.

We asked for teacher rankings (previous to SLEP testing) and compared with SLEP rankings. There was high coincidence, typically over 85 percent.

Wisconsin Public School.

Informally we use three factors to determine a student's placement (LEP level). I can say that there is generally a high relationship between SLEP scores and performance.

Preparatory School (Japan).

After having tested 200 students and worked with them for at least one academic year, I have noted a clear correlation between SLEP test scores and academic grades, later TOEFL scores, and oral English proficiency.

Other respondents (N = 16) who commented on SLEP's concurrent relationships with other measures, used language much like that cited above. For example:

(There is) direct correlation between SLEP scores and other ESL tests.

Students who score consistently higher on the SLEP are those who have relatively higher academic ratings and demonstrate a higher degree of oral proficiency.

(There is) fairly good correlation between SLEP scores and proficiency in written English, but not necessarily spoken English.

SLEP is a very accurate measurement.

SLEP reading scores reflect instructors' ratings of students' English reading skills.

There appears to be a loose correlation between SLEP results and academic teachers' ratings of ESL/EFL skills.

Relationships (with types of measures designated in the questionnaire) studied only informally, but the correlation is positive and seems high.

Based on informal observation, reading scores reflect class performance, but I wonder about whether high scores indicate readiness for 'academic reading' [from a U.S. university respondent who reported no formal study].

Teachers feel scores are good indicators of progress [from a U.S. high school ESL teacher who reported no formal study].

Several respondents commented on SLEP's role in placement.

International School (Switzerland).

SLEP so far shows to be highly accurate in enabling placement, with the proviso of placement affecting performance.

Intensive ESL Program (U.S.).

We have not made any formal study, but have found that we cannot rely solely on the SLEP for accurate placement. We have probably 10-15 percent of SLEP testers who are moved up or down following teachers' recommendations which disagree with SLEP results.

Independent Preparatory School (U.S.).

We have been able to make cut-off scores on the SLEP that are accurate as far as those students' ability to achieve in the class we assign them to.

Establishing readiness for English-medium instruction. Several respondents focused their comments on SLEP score levels at which students are judged to be ready to enter full-time English-medium academic programs, or indicated placement-levels (e.g., beginner, intermediate, advanced--not behaviorally defined in any instance) associated with specified SLEP score levels.

In the original SLEP validation sample, for ESL students who reported that they were in "mainstream classes" (full-time English-medium, academic instruction), the average SLEP total scaled score was 50 (see Exhibit A, above). It is noteworthy that several respondents who mentioned this factor independently identified SLEP scores at about this level as being indicative of readiness to enter English-medium academic programs. More specifically:

American Liberal Arts College (Japan).

It appears from our experience that students who score 48-50 have the ability to communicate in English in a way that would allow them to do academic work for credit. 51-60 usually means that their writing skills also are of a high enough level to engage in academic writing.

Canadian College.

ESL 3, 42-54; Mainstream 55 plus.

International School (Singapore).

Beginner (20-34), Low Intermediate (35-39), High Intermediate (40-47), Advanced (48-54).

Public High School (U.S.).

Students who generally get 48 and over are generally doing well and functioning in regular classes. They have a high transfer of skills from the primary language.

Independent School (U.S.).

We use a scaled score of at least 50 for placement into a regular English class.

Student Exchange Program A (U.S.).

We know that below 50 is a risk in one of the . . . member schools, and the score must be balanced by high results on other factors.

Canadian Continuing Education Program.

SLEP is used to make the general distinction between ESL and high school. We use 55 (raw score) on Listening and 50 (raw score) on Reading Comprehension as an average benchmark to admit students to high school. (The Form 1 total scaled score equivalent is 48).

Student Exchange Program B (U.S.).

Minimum scaled score for acceptance next year is 50.

Other respondents simply indicated that having relatively high SLEP scores was important to successful performance in English-medium programs--they did not cite clearly interpretable score levels. For example:

College-Related Preparatory School (Canada).

> *It is clear that a student in the 85-90 percentile is ready for the regular English program at our school and can integrate* **confidently** *into it.* (Reports initial means in the 47-53 range, SLEP reference group: 60th to 75th percentiles).

Exchange Program C (U.S.).

> *We have found that those students who far surpass the minimum score (not indicated) set for acceptance in the program usually have little difficulty functioning in an American high school.* (Typical student is in the 40-46 range when initially tested).

The comments of respondents, on balance, warrant the following general conclusions:

> SLEP scores were positively related to other measures of proficiency, including direct measures of oral English proficiency, essay ratings, teacher's ratings, and so on.
>
> SLEP scores provide generally acceptable (useful, accurate) bases for placing students according to proficiency levels, with the usual provision for adjustment in placement, based on actual performance in classes at the initial placement level.
>
> SLEP scores also have proven to be useful for screening prospective participants in exchange programs involving selection of students aspiring to study in English-medium preparatory schools.

Users' Perceptions of SLEP's Strengths and Limitations

Test users were invited to indicate what they perceived to be positive and/or negative aspects of the SLEP and the *SLEP Test Manual* and to make suggestions for improvement, using Questions Q8 through Q10, as indicated below:

> **Q8.** What do you regard as the most positive features of the SLEP Test (considering the uses indicated above)?
>
> **Q9.** And what are its primary limitations, from the same perspective?

Q10. What do you regard as the most positive/negative features of the *SLEP Test Manual*? What changes and/or additions to the Manual or the test itself would be most helpful to you?

Generally, comments regarding SLEP's most positive features emphasized various aspects of administrative convenience (ease of administration, availability of self-scoring answer sheets, and so on). Also emphasized were the fact that the SLEP is a standardized test of ESL proficiency that can be administered and scored locally, and that SLEP is valid (useful, helpful, accurate) for intended uses, and provides measures of both listening comprehension (sometimes referred to as "oral") skills and reading skills. Some respondents singled out the listening comprehension section for positive comment, while others (fewer in number) were especially impressed by the reading comprehension section.

Comments regarding perceived limitations of the SLEP, on the other hand, are less amenable to general summarization than those regarding positive features of the test. Whereas most of the positive comments pertained to identifiable aspects of the SLEP test itself (e.g., ease of scoring; validity or usefulness for local purposes, based on experience), many of the limitations mentioned were not "SLEP specific." Rather, they appear to be generalizable to any ESL proficiency test that only provides measures of listening and reading skills or to ESL proficiency assessment generally.

The comments of one ESL teacher in a public high school situation serve not only to highlight the recurring theme in these comments but also to suggest a logical, albeit difficult way to overcome the limitations involved:

ESL Teacher, U.S. High School.

> *(There) is no testing of oral language (writing skills). I would like to see a SLEP test that would include all four skills: listening, speaking, reading, and writing; easy to administer whether to one student or many at the same time.*

Variations on this theme are discernible in several brief excerpts from more extended comments.

> *SLEP does not measure student's ability to use the language directly; . . . has no writing/grammar component--must be supplemented; more (needed) on grammar and usage; . . . (we) need a test dealing with a broader range of skills for accurate assessment of progress; the SLEP does not measure proficiency in producing language (e.g., writing, speaking); . . . (it) only includes reading and*

listening comprehension; . . . (we) need a writing documentation since SLEP tests only LC and reading; (only a) limited range of knowledge (is) tested for; . . . (SLEP) does not measure speaking ability; . . . doesn't test written discourse; (the SLEP) does not assess production.

A smaller number of respondents indicated that high performance on the SLEP does not necessarily indicate a comparably high level of functional ability to deal with "academic English" in the classroom or in tests.

An ESL teacher offered the following comment, as well as pertinent interpretive insight:

ESL Teacher.

My students must pass a standardized English reading test at the 40th percentile; they often reach 90 percent on the SLEP but still are only about 20 percent-35 percent on English test. This may not be a limitation of the SLEP, but may deal with expectations of N. Y. State and this standardized English test.[16]

A few others commented more generally on the foregoing, typically observed pattern in the field of ESL proficiency assessment, as follows:

The SLEP doesn't test 'academic reading' ability; high-scorers may not be able to perform well academically.

The comments on perceived limitations of the SLEP, focus attention on the complexity of the assessment problems that confront ESL practitioners in SLEP-use contexts.

Other indicated limitations and/or suggested changes in the SLEP and/or the *SLEP Test Manual* call attention to specific modifications that are worthy of consideration on their merit, without regard to frequency of mention. For example:

- Include normal curve equivalent (NCE) conversions of percentiles in the *SLEP Test Manual*.[17]
- Provide a separately scored vocabulary section.
- Offer up-to-date norms; norms for specific subgroups.
- Additional forms would be useful.

- Provide more assessment of grammar/usage.

- Provide a taxonomy of item types according to the specific linguistic skills they are designed to assess (to enhance the usefulness of SLEP for diagnosis or for a more specific, curriculum-linked assessment of change).[18]

As noted earlier, the listening comprehension "map" items were negatively mentioned by seven respondents. Said one respondent,

> *In order to do well on these items, not only did one have to have good auditory memory, but also good spatial memory. I had difficulty with it, as I have poor spatial orientation.*

Other comments about the "map" items were less generically critical. For example:

> *(These items are) too difficult; . . . almost impossible for most students; . . . extremely difficult because of the inference that has to be done--i.e., where to buy a magazine. Or, map confusing, can't tell front of car from back easily.*

Of course, it does not necessarily follow from these comments that the "map" items are less valid than other item types in the SLEP. These comments indicate only that attention to both format- and validity-related questions regarding these items appears to be warranted.

The comments and suggestions by SLEP users, summarized above, point out potentially important general directions for further development and/or modification of the SLEP Test and/or the *SLEP Test Manual*.

The "Ideal Test Package"

> "You have commented on aspects of the Secondary Level English Proficiency test, and related matters. More generally, please describe briefly the characteristics of a standardized test of ESL/EFL proficiency (and related developer-provided materials/services) that would be most helpful/ useful in EFL/ESL assessment contexts similar to yours."

Nineteen respondents provided comments and/or suggestions regarding an idealized ESL "assessment package." As might be expected from many of the comments on "limitations" reviewed above, a recurring theme was that the "ideal" test battery would provide for assessment of all four basic skills and offer enhanced diagnostic potential.

The potential usefulness of a "lower level" test was noted by two respondents. Some respondents wanted a somewhat shorter test, with features designed to facilitate its administration. Others wanted a test that was free of "cultural bias" and "gender bias."

The detailed comments providing the basis for the foregoing summarization are included below. In several, less detailed comments, some practitioners petitioned for a breakdown (of information regarding test items) like that done by the publishers of CTBS tests (a U.S. achievement test battery), updated data on the relationship between SLEP and the TOEFL, or a measure of ability to "use academic language."

One teacher called attention to the complexity of assessment involving "third-world" students with limited academic preparation, while another (from an independent secondary school in the U.S.) indicated simply that the "SLEP seems to provide most of what is needed at this particular school."

More comprehensive assessment

Testing Director, Japanese Branch U.S. University.

> *For the purposes of placement, a more comprehensive standardized test would be welcome, i.e., one which includes balanced components measuring writing ability and speech production in addition to listening and reading comprehension. Given the homogeneous nature of our particular EFL context, however, I believe that a test developed for this particular population might be more useful. It would be difficult for me to describe a standardized test that would be more appropriate.*

ESL Department Head (Canada).

> *In addition to the listening and reading skills, some organized way of measuring speech and writing (would be helpful). It must be something that can be administered without a heavy commitment of instructor time.*

English Department Chairperson, Academy (U.S.).

> *A thorough assessment of a student's proficiency in English usage, grammar, speaking, reading, etc. Content and context geared to high school students' interests, experiences. Test which is easily scored.*

Director of Testing, Preparatory School (U.S.).

I want to see vocabulary (in context) strength (or weakness), grammar knowledge, (prepositions, verb usage). Idioms are not important at our level--the students pick these up in class. Oral expression might be assessed by audio-tape. For admissions purposes, we cannot handle students who cannot make themselves understood at a primary level.

ESL Teacher, Migrant Education (U.S.).

I would like to see a standardized test that would tell me the areas of weakness. Something that would be helpful to teachers, so they could zero in on the areas of weakness and provide practice and language instruction that would improve these areas.

ESL Dept, International School (Japan).

I would like to see a production component and some consistent assessment of production added or available as a supplement.

Reduce testing time

Head, ESL Program Canada (College-related School).

A 45 to 60 minute listening/reading comprehension test which requires no introduction by the teacher and which has introductory instructions in many different languages so that the student can begin with confidence.

ESL Coordinator (U.S. High School).

A test that included the testing of the 4 skills--listening, speaking, reading, writing. A test that takes less time to administer and grade. It is difficult **in my program format** *to administer SLEP as a determining factor for ESL program entry (emphasis added).*

A "lower-level" test

EFL/ESL Specialist, School System (U.S.).

It would be most helpful if a reading comprehension test at a lower level was available. Many of our refugee students have little or no education. I would like to be able to assess their skills better. However, this is a good test, and I intend to continue to use it.

English Director (U.S. Public School).

A second language test for students in lower grades would be helpful.

Freedom from cultural and gender bias

Executive, American School (Europe).

Ours is a unique program--not placement alone--but personal qualities must be assessed. SLEP serves its purposes but would never serve alone. Only addition would be writing sample. I assume research is done in regard to gender and cultural background being unbiased. This would be very important--that SLEP test is not biased to sex and to American cultural background (but probably unavoidable).

Executive, International College (Japan).

Tests without a lot of culturally biased vocabulary and subject matter. A weighted test which can easily be used to level or sequence students and curriculum needs.

The opportunity for humor lurking in an invitation couched in such a way as to suggest the possibility of devising a test that would meet the extremely complex assessment demands that confront ESL practitioners, was seized upon by one respondent who characterized the ideal test as follows:

Supervisor, ESL/Bilingual Program (U.S.).

The ideal test would serve well for both student assessment and program evaluation. It would be a criterion-referenced test (magically based on our own curriculum) that can also be interpreted by norms. Information would be sufficiently rich to yield placement, diagnostic, and normative interpretations. In addition, such a test would provide data on growth, gains, and gap reduction that would satisfy federal reporting requirements and our own omnivorous curiosity.

Summary

The findings of this survey represent the results of an ad hoc, formal effort to obtain feedback from practitioners in diverse SLEP-use contexts. Although small, the sample of respondents appears to be generally representative of the sample surveyed with respect to type of institution and location (U.S.A, Canada, other country). General trends in findings are summarized below.

- Almost 90 percent of the respondents reported using SLEP Form 1, 70 percent reported using Form 2, and 37 percent reported using Form 3.

- About one-third of the respondents indicated that the SLEP was the sole ESL proficiency measure being used.

- The number of examinees tested annually varies considerably across use contexts; about 34 percent reported testing less than 50 examinees, and 19 percent reported testing 250 or more examinees.

- In the majority of cases (60 percent) examinees are tested at least two times.

- Slightly more than one-half (52 percent) of the sample reported that testing was restricted to students in the G7-12 range; more than one-third reported testing college-level students; some 20 percent reported testing sixth graders. Respondents' comments indicate that the SLEP is perceived to be generally suitable for use with examinees at quite diverse educational levels ranging upward from sixth grade through college--but may tend to be relatively more suitable for college-level than for sixth-grade level examinees.

- Respondents from use contexts in the U.S.A. and Canada reported local populations comprising not only "international students" (typically accounting for about 43 percent of examinees), but also resident aliens, recent immigrants, refugees, undocumented individuals, and so on. In other countries, testing populations comprised primarily local residents and other nonnative-English speakers studying or planning to study in English-medium preparatory schools or colleges, situated locally or elsewhere.

- Local examinee populations differ rather markedly in heterogeneity of language background. In about one-third of the settings, only one language group is represented; in some 29 percent of the settings, eight or more language groups are represented.

• The SLEP is being used, typically, for at least two of the purposes that are recommended in the Manual--that is, to assess readiness for undertaking full-time English-medium academic programs (in about 70 percent of use contexts), to assess average gain (estimated 50 percent) for ESL placement (about 45 percent), or in screening for admission to institutions or programs (about 33 percent). In addition, some 57 percent of U.S.A./Canada respondents and about 31 percent of all others indicated use of the SLEP for monitoring the progress of individual students--a practice not specifically mentioned in the *Manual*. It was not described in detail by any of the respondents reporting it.[19]

• Slightly more than one-half of the sample (58 percent) indicated they had conducted local studies of the relationship between SLEP scores and other measures (e.g., teacher's ratings of oral English proficiency). Only 14 percent reported they had developed "local norms for the SLEP (e.g., a table showing the percentage of students scoring at or below designated SLEP scores)." By inference from the nature of the comments provided by respondents, some local assessments of SLEP's validity and usefulness are quite sophisticated, but most of them are relatively informal--frequently involving primarily clinical perception rather than statistical documentation.

• At the same time, there was a relatively consistent "positive validity" theme in the comments. Respondents relatively consistently reported having observed positive relationships between SLEP scores and more direct measures, such as those referred to in the question (see Exhibit A, above). They often indicated generally that the SLEP had been found to be "valid" or "useful" for local purposes.

• Respondents named as "positive features of the SLEP Test," its administrative convenience, the fact that it is a standardized test of both listening and reading skills, and its validity/usefulness for local purposes.

• Comments on "negative features" of the SLEP typically did not single out for criticism any specific features of the SLEP. Rather, the recurring theme reflected a need for a more comprehensive measure. More specifically, the fact that the "SLEP does not test production" was mentioned with relative frequency as a limitation of the SLEP. Several respondents also indicated, as a negative feature of the SLEP, that high performance on the SLEP does not necessarily indicate a

comparably high level of functional ability to deal with "academic English" in classroom settings--clearly a generic problem.

- Nineteen respondents accepted the invitation to describe a "test package" that would be most helpful/useful to them. Consistent with the general comments on SLEP's "limitations," a recurring theme was that the ideal test battery would provide for assessment of all four basic skills and offer enhanced diagnostic potential. Less prevalent themes called for reducing testing time by developing a shorter test, a test for lower grade levels, and a test that is free from gender or cultural bias.

Some of the implications of these findings for research and development (R&D) activities involving the SLEP are discussed in the following section.

Implications of the Survey Findings

The information, ideas, comments, and suggestions of survey respondents are useful and important, on merit, without regard to statistical considerations generally or to the fact that only a small percentage of the total population of SLEP-users responded to the survey questionnaire. The responding sample, as indicated, appears to be representative of the general test-using population. The fact that the SLEP is being used relatively extensively with postsecondary-level students is noteworthy; more than one-third of the respondents reported that the SLEP was being used with college-level students.

Based on respondents' descriptions of findings of local studies and/or their clinical observations, scores on the test have been found to be positively related to other indices of ESL proficiency, including direct assessments of oral English proficiency and writing skills, across samples from diverse test-use contexts.

This feedback and other pertinent evidence[20] suggests as a strong working hypothesis that the SLEP can be expected to provide reliable and valid measurement of ESL listening comprehension and reading skills in samples of college-level students as well as in samples of younger students. Thus SLEP's identification as a test designed for use with "secondary level" students appears to be unduly restrictive in its connotations.

Additional research is needed, however, to establish SLEP's validity in an expanded population, to extend evidence of validity generally, and to develop up-to-date and comprehensive reference-group data for samples from representative, current, and potential SLEP-use contexts.

Generally speaking, formal research-based evidence bearing on SLEP's reliability and validity is relatively limited--as compared to the large body of evidence bearing on TOEFL's validity, for example. The only reference group available reflects the performance of ESL students in grades 7 through 12 in approximately 50 U.S. public schools, tested circa 1980 (with SLEP Form 1). The need for updated and expanded reference group data for the SLEP was specifically noted by a number of respondents and is recognized by the SLEP School Services Program.

The development of current, comprehensive reference-group data for both secondary-level and postsecondary-level samples, classified by gender, language background, and other pertinent variables, is needed to enhance the usefulness of the SLEP (as well as to establish or maintain its "certifiability" for use in certain contexts).[21]

Local SLEP users, in postsecondary-level and high-school level settings, are in a position to contribute directly to the development of reference-group data and additional formal evidence bearing on SLEP's validity by participating in cooperative studies designed to collect SLEP scores, background data, and pertinent criterion data from representative testing contexts.

Selected SLEP users might be invited to provide SLEP data and ratings or scores on a "common criterion measure" (e.g., grades in ESL courses, ESL teacher's ratings of proficiency according to a standard scale, and so on) for defined samples. Given such data, it would be possible to conduct centrally the types of analyses needed to assess the strength and consistency of association between SLEP scores and the criterion measure(s) involved.

Because the TOEFL is widely used and has been extensively validated for postsecondary-level samples, it would be useful to conduct studies designed to extend evidence regarding the strength and consistency of SLEP/TOEFL relationships in postsecondary-level samples. Similar studies should be conducted in samples of secondary level students in settings where attaining levels of proficiency indexed by TOEFL scores represents an important goal for the students involved.[22]

Other Avenues for SLEP-Related R&D Activities

As a measure specifically designed to gauge listening comprehension and reading skills, the SLEP obviously cannot meet the complex range of assessment needs and concerns expressed by the respondents to this survey--including the need for comprehensive assessment of productive as well as receptive skills.[23]

Promote widespread use of standard procedures for rating productive skills

It is important to promote the local use of standard procedures for assessing writing and speaking skills. For example, consideration might be given to the development of brief, behaviorally anchored rating schedules that ESL teachers could use in rating essays or speaking ability--perhaps adaptations of currently available scales for evaluating these skills.[24] In any event, it seems important to encourage SLEP users to adopt standard procedures for rating basic skills--procedures whose usefulness could be explored in cooperative studies in which the ratings constitute "common criteria" across use contexts.

Explore SLEP's validity below the G7-12 range

Survey findings indicate that the SLEP is being used for sixth graders in a number of contexts. Evidence bearing on SLEP's "suitability for use with 6th graders," is quite limited. Some respondents suggested the potential usefulness of a "lower level" of the SLEP for use below the G7-12 range. Research is needed to assess SLEP's difficulty, reliability, and validity in samples below the 7th grade level.

Available evidence (e.g., ETS, 1991; Holloway, 1984) suggests that most native English-speaking seventh graders have "mastered" the skills measured by the SLEP--that is, they tend to "top out" on the SLEP, more so on listening comprehension than on reading. At what age/grade level do SLEP items begin to represent relatively difficult cognitive tasks for native English-speaking students? A study designed to answer this question would provide information that is pertinent to the problem of establishing the lower "age/grade limit" of SLEP's applicability.

Increase SLEP's "assessment efficiency"

The amount of time required to administer the SLEP--the amount of time needed for placement testing generally--was a matter of some importance for a number of the ESL practitioners who responded to the survey, as it was to those interviewed a decade earlier by Hale and Hinofotis (1981: pp. 10-11). It

seems important to consider research and development activities designed to explore options that might result in "increased efficiency of measurement," for assessments involving the SLEP test--for example, by reducing testing time, and by introducing features that would capitalize on the general (class-level) diagnostic potential inherent in SLEP items.[25]

Explore the reliability and validity of a shorter test

Time needed for ESL proficiency assessment is a matter of considerable importance in SLEP-use contexts. The SLEP, as presently constituted, requires approximately one and one/half hours of testing time. In one major testing context involving member institutions of the Los Angeles Community College District (LACLCD), the colleges use an "abbreviated" version of the SLEP for the express purpose of reducing the total amount of time needed for placement testing.

The LACCD reduces SLEP-testing time simply by not administering two of the sections (Tillberg, 1991, personal communication). This approach to reducing testing time and the particular item types selected for inclusion/exclusion were, as recommended by Butler (1989), based on analyses that included an assessment of the comparative validity of scores on the full and shortened versions for discriminating among independently defined ESL proficiency-placement groups.

It would be useful to conduct research designed to assess the effects of the approach described above and other approaches to reducing the length (and time required for administration) of the SLEP test, on reliability, concurrent validity, validity for placement and other specific purposes, and so on. Exploratory research might be conducted, retrospectively, using existing data sets that include item-level scores for the complete SLEP test and criterion scores (teacher's ratings, and so on).

Assess contribution of item types to validity

Little attention has been given to assessing the comparative validity of the respective SLEP item types for predicting basic performance criteria (e.g., ratings of oral language proficiency or writing ability).[26] Studies of the relative validity of SLEP item types would contribute information that is relevant to the problem of developing a shorter test. The studies, incidentally, would also contribute to an empirical evaluation of the validity-related properties of the map items that were mentioned negatively by several respondents. In this same general area, it would be useful to analyze the factor structure of the SLEP using, for example, data sets supplied by SLEP users.

Enhance SLEP's general diagnostic potential

The SLEP was developed to provide a reliable and valid basis for assessing ESL listening comprehension and reading comprehension. Attention naturally is focused almost exclusively on the reliability, validity, and usefulness of these two scores and the total score. Little attention is paid to the variety of subskills that may be tapped by different sets of test items. These are of potential interest to ESL teachers and others interested in identifying general instructional areas that may need more or less attention in plans for instruction.

According to one ESL/Bilingual supervisor, for example:

> *Although the SLEP is not a criterion referenced test, it would be helpful to know what underlying skills or curriculum goals, if any, are addressed by the test items.*

Developing a taxonomy of skills/functions tapped by existing SLEP items would contribute directly to increased efficiency of assessment by enhancing the general diagnostic potential of the test. Even though the SLEP is not designed with particular curricular goals or discrete skill development in mind, the types of skills/function represented in the test items are likely to be common foci of instruction in most ESL curricula. Averages based on subsets of items by skill areas appear to have potential value for general evaluation and instructional purposes.[27]

Cultivate "Cooperative Interaction" with SLEP Users

It would be useful to consider procedures designed to promote closer ties and more frequent professional and collegial interaction between the SLEP School Services Program and the ESL practitioners who administer and use the SLEP in diverse local settings, worldwide.

An important, albeit simple, step in that direction would be to modify SLEP ordering procedures by asking for full professional and personal identification of the "individual who will be responsible for using SLEP." This information is essential to the definition of a "population of SLEP users," as well as "SLEP ordering institutions." Lack of personal identification for SLEP users complicates efforts to interact with those who are actually using the test, as indicated by the difficulties the present survey encountered in identifying and contacting "SLEP users", outlined at the outset.

Other steps that might be considered include establishment of a "SLEP Advisory Service," including a toll-free "hotline" through which practitioners can raise and receive answers to questions about SLEP use and interpretation. A periodic "newsletter" would provide a means of keeping test users informed of developments regarding SLEP. If SLEP users were encouraged to provide reports of local studies, these results could be shared periodically with all test users through the newsletter, and so on.

Steps taken to encourage and facilitate professional interaction between the SLEP School Services Program and SLEP users should be beneficial to all involved. Consideration might be given to the development of a model for implementing a program of cooperative interaction between the School Services Program and SLEP users, that would involve periodic data feedback from test users in exchange for central analysis and reporting by the program.

Appendix A: Illustrative SLEP Items

▪▪▪▪▪▪▪ Sample Questions

Section 1

The first section of the SLEP test measures ability to understand spoken English and is 35-40 minutes long. It is divided into four parts, with four different types of questions.

Part A

For the first type of question, the student must match one of four recorded sentences with a picture in the test book The sentences are spoken only once and are not printed in the test book. This part contains items dealing with correct recognition of minimal pair contrasts, juncture, stress, sound clusters, tense, voice, prepositions, and vocabulary.

Sample Questions

Note: Pictures are for illustrative purposes only. Actual pictures and drawings in the test booklet are two to four times larger than sample pictures in this brochure.

1. On tape:
   Look at the picture marked 1.

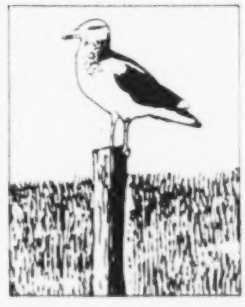
   On tape:
   (A) *There is an arrow in the sky.*
   (B) *The building has a tall tower.*
   (C) *The judge is bowing his head.*
   (D) *There is a toy in front of the building.*

1.

2. On tape:
   Look at the picture marked 2.

   On tape:
   (A) *The bird is standing on top of the pole.*
   (B) *The bird is flying over the fence.*
   (C) *The bird is digging in the sand.*
   (D) *The bird is eating the grass.*

2.



3. On tape:
Look at the picture marked 3.

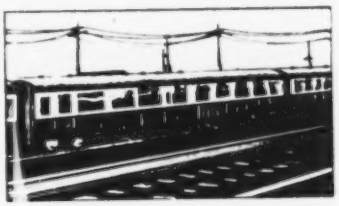
On tape:
(A) *There's a statue of a lion.*
(B) *The line is very straight.*
(C) *The wine is near the window.*
(D) *There's a lane near the building.*

3.

4. On tape:
Look at the picture marked 4.

On tape:
(A) *The brain is protected by bone.*
(B) *The train is on the track.*
(C) *The drain is stopped up.*
(D) *The rain is coming down.*

4.



Part B

These questions approximate the type of dictation exercises used frequently in English language classes: the student must match a sentence printed in the test book with a sentence heard on the tape. The questions focus on the relationship between structure and meaning.

Sample Questions

1. On tape: *The class can finish it in less than an hour.*

In test book: (A) The classes can't finish in half an hour.
(B) The class won't be finished for an hour.
(C) The classes will take at least an hour.
(D) The class can finish it in less than an hour.

2. On tape: *Why aren't they fixing the car?*

In test book: (A) Are they fixing the car?
(B) I'm fixing the car.
(C) Why aren't they fixing the car?
(D) The car has been fixed.

3. On tape: *While I was waiting for my sister, she got the news.*

In test book: (A) While I was waiting for my sister, she got the news.
(B) While my sister was waiting for me, she got the news.
(C) I was waiting for my sister to get the news.
(D) I was waiting for my sister when I got the news.

4. On tape: *He didn't know how to get to the gym.*

   In test book: (A) He didn't go to the gym.
   (B) He explained how to use the gym.
   (C) He told us to get to the gym.
   (D) He didn't know how to get to the gym.

5. On tape: *Bill has one brother and one sister, and so does Jane.*

   In test book: (A) Bill has one brother and one sister, and so does Jane.
   (B) Bill has one brother and a sister named Jane.
   (C) Bill and Jane are brother and sister.
   (D) Bill's brother and sister like to be with Jane.

Part C

For the second type of question, the student refers to a map in the test book (see page 11). Streets and buildings on the map are labeled, and there are four cars, marked A, B, C, and D. The student must choose the one car that is the source of a brief conversation on the recording. The questions in this part assess a variety of linguistic, cultural, and pragmatic concepts. These include directions, recognition of building names and associated vocabulary, distance, and time.

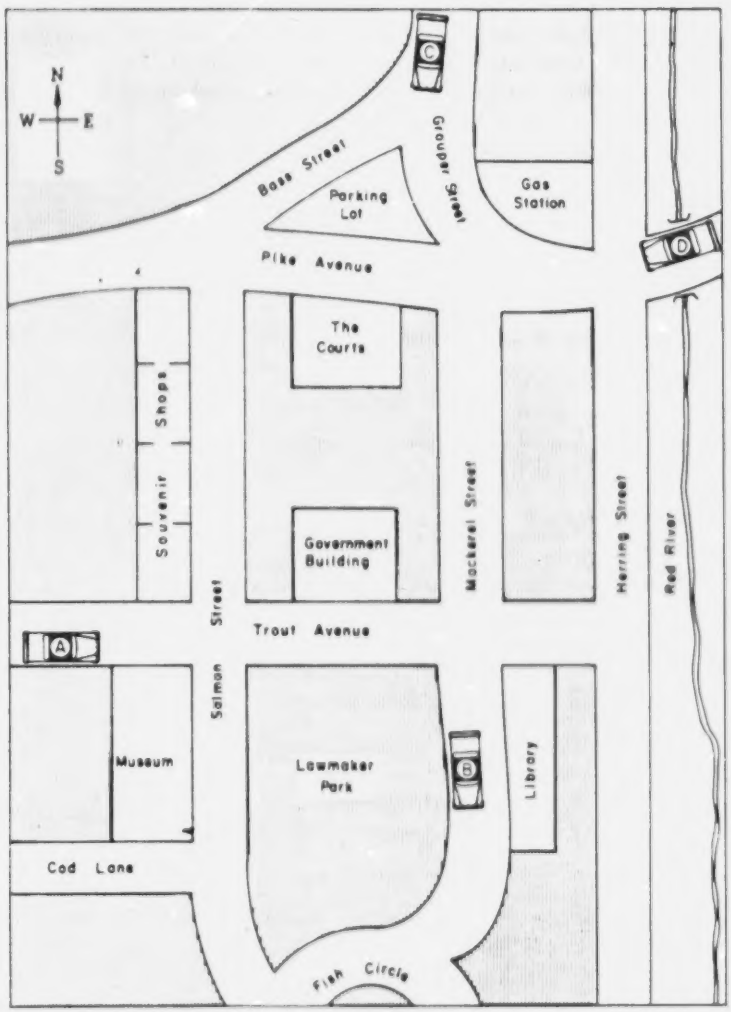
Sample Questions

1. On tape:
   (man) *The museum has a special exhibit this week. Why don't we go?*
   (woman) *I'd like to very much. If we continue on Mackerel to the circle and go around to Salmon, we can park on Cod Lane.*
   (third voice) *Which car are the people in?*

2. On tape:
   (man) *I would like to find the way to the circle. From there, I know how to get home.*
   (woman) *It's not too hard. If we bear right into Bass and then go south on Salmon, we will end up at the circle.*
   (third voice) *Which car are the people in?*

3. On tape:
   (woman) *The judges are going to hear a very interesting case today. Let's stop at the courts.*
   (man) *That's a good idea. I'll go north at the next intersection and cross Pike Avenue. We can park in the lot across the street from the courts.*
   (third voice) *Which car are the people in?*

Appendix A, con't: Illustrative SLEP Items



BEST COPY AVAILABLE

Part D

The questions in this part are based on conversations, recorded by American high school students, that represent typical secondary school situations. The conversations take place in various parts of a school and deal with events that typically occur in each location. Conversations also deal with extracurricular activities, academic subjects, school closings, and holidays. For each recorded question, the student must choose one of four answers printed in the test book.

Sample Questions

1. On tape:
   (Bob) *I heard that it is supposed to be a very good band. Since the game starts at 7:30, Nancy, I'll pick you up at 7.*
   (Nancy) *That's fine. I'll be ready. It takes 15 minutes to get to the gym, so we'll have time.*
   (third voice) *At what time will they arrive at the gym?*

   In test book: (A) 6:45.
   (B) 7:00.
   (C) 7:15.
   (D) 7:30.

For questions 2 and 3.

2. On tape:
   (Nancy) *Jane, what are you going to wear to the game?*
   (Jane) *I'm not sure yet. I don't want to have a heavy sweater on at the dance. It'll be pretty warm in the gym. I'll probably wear a light dress, even though the weather outside might not be so warm.*
   (third voice) *What is the girl going to wear?*

   In test book: (A) A heavy sweater.
   (B) A heavy coat.
   (C) Some light slacks.
   (D) A light dress.

3. (On tape) *What is the girl's reason for this decision?*

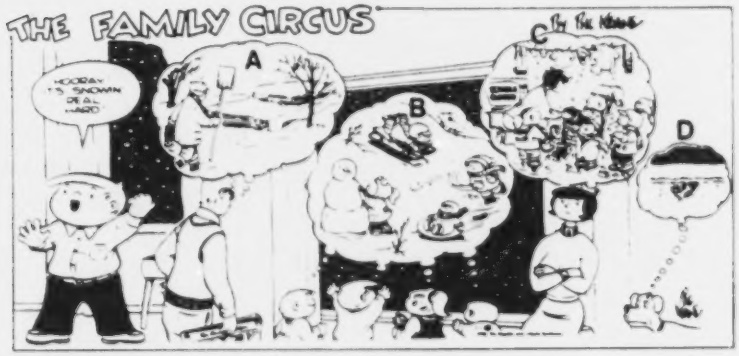
   In test book: (A) She expects it to be cold outside.
   (B) She expects it to be warm inside.
   (C) It is going to snow.
   (D) It will be very windy.

Section 2

The second section of the test is 40 minutes long and measures ability to understand Written English. The questions cover grammar, vocabulary, and reading comprehension. There are three parts to Section 2.

Part A

For each question in this part, the student must match the reaction of one of four characters in a cartoon with a printed sentence.



Sample Questions

1. All those wet clothes. The children will want to stay outside and I'll spend my time trying to keep them dry.

2. I can hardly wait to make the first snowball. I've been waiting all year to get back at her.

3. Oh, my aching back. The car will be covered and I'll have to shovel it out.

4. Isn't it great that school might be closed? I'd much rather have fun outside than stay in school. What better way to spend a snowy day.

5. I'm going to be awfully hungry. I shouldn't have hidden that bone. It would have been better to leave it in the house.

Part B

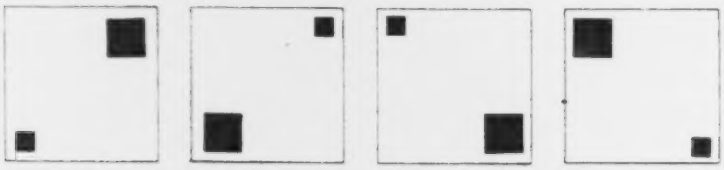
For the questions in this part, the student must match a printed sentence with one of four drawings. The particular focus of this item type is the use of prepositions, pronouns, adverbs, and numbers.

Appendix A, con't: Illustrative SLEP Items

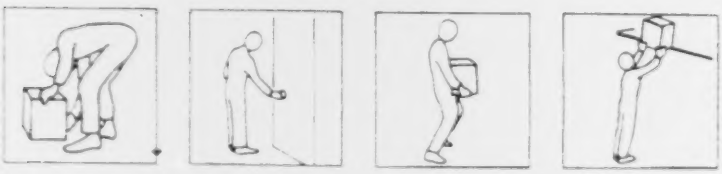
Sample Questions

1. One girl is eating ice cream but two aren't.

2. The small square is in the upper left corner.



3. He is bending over to pick up the box.



4. The car almost hit him while he was crossing the street.

Part C

This part of Section 2 contains questions of two types. In one, the student must complete passages by selecting the appropriate words or phrases from among four choices printed at intervals in the passages.

Sample Passage and Questions

1. Sound is something we (A) hears. (B) hearing. (C) heard. (D) hear. It comes to your

2. (A) eyes (B) nose (C) ears (D) mouth in different ways. It might be pleasant,

3. like the voice of a friend, (A) when (B) as (C) or (D) since unpleasant, like the screech

4. of a train's wheels on a railroad (A) station. (B) track. (C) light. (D) conductor. Some sounds are loud,

5. and some are soft; some are high, and some are (A) full. (B) low. (C) quiet. (D) big. Sound is

6. very (A) importance (B) importantly (C) important (D) import to us because it is the basic means of communication.

In the second type of question, the student must answer questions about the passage for which he or she supplied the missing words or phrases.

Sample Questions

7. What does screech in line 3 mean?

(A) noise (B) motion (C) place (D) piece

8. Which of the phrases below is another example of a pleasant sound, similar to the phrase in the sentence that begins in line 2, "like the voice of a friend"?

   (A) Like the ring of an alarm (B) Like the wail of a siren
   (C) Like the honk of a horn (D) Like the song of a bird

9. Which sentence below has almost the same meaning as the sentence that begins in line 5?

   (A) It is meaningful to communicate with sound.
   (B) The main way we communicate is with sound.
   (C) The meaning of sound is basic to communication.
   (D) In order to communicate, we need basic sounds.

Part D

In this part of Section 2, the student must read a short literary passage and answer questions about it.

Sample Passage and Questions

> The footsteps began about a quarter past one o'clock in the morning, a rhythmic, quick-cadenced walking around the dining room table. My mother was asleep in one room upstairs, my brother Herman in another; grandfather was in the attic, in the old walnut bed. I had just stepped out of the bathtub and was busily rubbing myself with a towel when I heard the steps. They were the steps of a man walking rapidly around the dining room table downstairs.

1. What did the writer hear?

   (A) A soldier marching (B) His brother snoring
   (C) His mother talking (D) A person walking

2. Where did the sounds come from?

   (A) The attic (B) The dining room
   (C) The bathroom (D) The stairs

3. What was most of the family doing?

   (A) Listening (B) Working (C) Bathing (D) Sleeping

Appendix A, concluded: Illustrative SLEP Items

4. What was the writer doing?

   (A) Talking to himself (B) Drying himself
   (C) Brushing his hair (D) Getting dressed

5. The bed in the attic was made of which of the following materials?

   (A) Metal (B) Wood (C) Feathers (D) Straw

6. What time did the sounds begin?

   (A) 12:45 p.m. (B) 1:00 a.m. (C) 1:15 a.m. (D) 1:30 p.m.

Appendix B: The Survey Questionnaires

Exhibit B.1: Cover letter and general question

Exhibit B.2: Questionnaire for U.S. and Canadian test users

Exhibit B.3: Questionnaire for test users in all other countries

Exhibit B.1: Cover Letter, General Question

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 085-

April 17, 1991

Dear Colleague:

Since 1980, the Secondary Level English Proficiency (SLEP) test, developed by Educational Testing Service (ETS) for assessing the English-language listening comprehension and reading skills of nonnative-English speaking (ESL/EFL) students in the G7-12 age/grade range, has been administered and scored locally in scattered settings throughout the world. The SLEP program needs, but does not regularly receive, feedback from test users regarding the variety of purposes for which the test is being used, the age/grade levels and language backgrounds of the students being tested, perceived strengths and limitations of the test for particular purposes, and so on. Without such feedback, the program is not in a position to judge the extent to which current forms of the SLEP are meeting the needs of users and introduce modifications designed to improve the overall usefulness of the test from the perspective of practitioners in diverse use settings.

By inference from information supplied by the SLEP testing program regarding orders for SLEP booklets and related materials in recent months, it appears that the SLEP is being used or considered for possible use in one or more programs in your setting. The brief questionnaire enclosed is designed to obtain feedback regarding the types of issues indicated above. Survey findings will be summarized statistically, and survey respondents will receive a brief summary report in which neither individual respondents nor their institutions will be identified directly with particular findings. Respondent identification, called for on the cover of the questionnaire, is needed to facilitate followup inquiries that may be needed to clarify particular questionnaire responses and to identify the individuals most directly concerned with use of the SLEP examination (to whom copies of the survey summary will be sent). A prepaid business-reply envelope is enclosed for returning the completed questionnaire.

Your assistance in completing and returning the questionnaire, or in forwarding this letter and the enclosures to the individual who is most directly involved with SLEP use in your setting, will be greatly appreciated.

Sincerely,

Kenneth M. Wilson
Research Psychologist

Copy for: Ms. Stella Cowell
Director, SLEP Program

Encl: Questionnaire and return envelope

FROM A PRACTITIONER'S PERSPECTIVE

You have commented on aspects of the Secondary Level English Proficiency test, and related matters. More generally, please describe briefly the characteristics of a standardized test of ESL/EFL proficiency (and related developer-provided materials/services) that would be most helpful/useful in EFL/ESL assessment contexts similar to yours?

BEST COPY AVAILABLE

THE SECONDARY LEVEL ENGLISH PROFICIENCY TEST:
A Survey of Test Users

Person to whom further inquiries about this survey should be addressed:

Name: _____
Title: _____
Institution: _____
_____
_____

Telephone number: _____ FAX number: _____

.....

Context in which SLEP is being used:

1. ESL/EFL proficiency assessment in the context of ongoing academic programs
2. ESL/EFL proficiency assessment in an institution specializing in language instruction (no academic program)
3. ESL/EFL proficiency assessment in a business or corporate setting
4. ESL/EFL proficiency assessment by or on behalf of an agency concerned with facilitating the recruitment and/or placement of international students in English-medium academic programs
5. Other assessment context (please specify)

Please provide a brief elaboration of the response checked above (the type of setting in which ESL/EFL assessment is being conducted, the goals of ESL/EFL assessment, etc.):

_____
_____
_____
_____
_____

.....

If you have questions about the survey:

Contact Kenneth M. Wilson, 12-R, Educational Testing Service, Princeton NJ 08540 (USA)

Telephone: 609 734 5391
FAX No.: 609 734 1090

In answering the following questions regarding use of the SLEP test, please consider the period: Spring, 1990 to Spring, 1991.

Q1. The SLEP is being used for (check all applicable and elaborate briefly below):

1. ___ **Assessing readiness of ESL/EFL students to undertake full-time English-medium academic programs**
2. ___ **Deciding which of several alternative programs (e.g., full-time vs. part-time ESL, bilingual) is most appropriate, based on observed level of ESL/EFL proficiency**
3. ___ **Assessing average gain (using mandated evaluation models) in tested ESL/EFL proficiency for students in programs defined/required by statute (Federal/State/Provincial)**
4. ___ **Assessing average gain for students in ESL/EFL, bilingual, or similar programs generally, using locally devised test-retest procedures**
5. ___ **Screening applicants for admission to an academic program/institution**
6. ___ **Monitoring the progress of individual students**
7. ___ **Other (specify):**

Please elaborate briefly:

_____
_____
_____
_____
_____

Q2. Forms of the SLEP used (check all applicable):

1. ___ Form 1 2. ___ Form 2 3. ___ Form 3

Q3. ___ Please enter in the preceding space the approximate number of individuals tested with the SLEP within the past 12 months.

Q3a. Indicate how many times the typical individual was tested: ___ 1. Once ___ 2. Twice

Q4. Estimated percentage of individuals tested in each of the following categories:

___ 1 Below 7th grade level (please answer Q4a)
___ 2 Grades 7 - 12
___ 3 College level or beyond (please answer Q4a)

Q4a. Please indicate age/grade levels involved, and comment on the extent to which the SLEP appears to be suitable/unsuitable for such younger/older students, giving specific examples.

_____
_____
_____
_____
_____

BEST COPY AVAILABLE

Q5a. Which of the following patterns of test use is most applicable for your setting?

1 The SLEP is the only standardized measure of English proficiency used (for the purposes indicated above)
2 The SLEP and locally developed tests or assessment procedures are being used
3 The SLEP, other standardized ESL/EFL tests, and locally developed tests or proficiency assessment procedures are used

Q5b. Please describe briefly typical testing practices indicating how the SLEP, other standardized tests, and/or locally developed tests are used to accomplish overall objectives.

Q5c. Please write in the names of up to three of the regularly or frequently used additional English proficiency tests.

Q5c. Have you been able to study the relationship between SLEP scores and direct measures of ESL/EFL proficiency (for example, ESL/EFL instructor's ratings of oral English proficiency; academic teachers' ratings of students' ESL/EFL skills)?

1 Yes (please describe briefly)
2 No (please comment briefly on your impressions regarding the foregoing, and reasons for them)

Q5d. Five SLEP Total scaled score categories are specified below. Please check the score range that includes the average score typically obtained by students when initially tested.

1. < 32 2. 33-39 3. 40-46 4. 47-53 5. 54+

Q5e. Have you developed local norms for the SLEP (e.g., a table showing the percentage of students scoring at or below designated SLEP scores)?

1 Yes (If possible, please enclose a copy of your norms table and related description)
2 No

Q6. Approximately what percentage of all students tested are classifiable as: (Total should be 100%)

1 ____ Permanent residents for whom English is a second language
2 ____ Recent immigrants
3 ____ Refugees
4 ____ Undocumented
5 ____ International students (in or planning to enter an English medium secondary- or postsecondary-level program)
6 ____ Other (please describe)

Q7. Several general language groups are listed below. Please check all groups represented in your testing population. Please write in language groups that account for a higher proportion of the testing population than any of those listed.

1	Spanish	9	Thai
2	Japanese	9	Portugese
3	Chinese	10	Indonesian
4	Arabic	11	French
5	Korean	12	
6	Cambodian	13	
7	Vietnamese	14	

Q8. What do you regard as the most positive features of the SLEP test (considering the uses indicated above)?

Q9. And what are its primary limitations, from the same perspective?

Q10. What do you regard as the most positive/negative features of the SLEP Test and Score Manual? What changes and/or additions to the Manual or the test would be most helpful to you?

See back of cover memorandum for a general question about ESL/EFL tests. Thanks for your help.

BEST COPY AVAILABLE

THE SECONDARY LEVEL ENGLISH PROFICIENCY TEST

A Survey of Test Users

Person to whom further inquiries about this survey should be addressed:

Name _____
Title _____
Institution _____
_____
_____
_____
_____

Telephone number _____ FAX number _____

Context in which SLEP is being used:

1. ESL/EFL proficiency assessment in the context of ongoing academic programs
2. ESL/EFL proficiency assessment in an institution specializing in language instruction (no academic program)
3. ESL/EFL proficiency assessment in a business or corporate setting
4. ESL/EFL proficiency assessment by or on behalf of an agency concerned with facilitating the recruitment and/or placement of international students in English-medium academic programs
5. Other assessment context

Please provide a brief elaboration of the response checked above (the type of setting in which ESL/EFL assessment is being conducted, the goals of ESL/EFL assessment, etc.):

_____
_____
_____
_____
_____

If you have questions about the survey:

Contact Kenneth M. Wilson, ETS, Educational Testing Service, Princeton, NJ 08540 (USA)

Telephone: 609/734-5291
FAX No.: 609/734-1090

In answering questions regarding use of the SLEP test, please consider the period, Spring, 1990 to Spring, 1991.

Q1. The SLEP is used for (check all applicable):

1 ___ Assessing readiness of students to undertake full-time English-medium academic programs

2 ___ Making decisions regarding the placement level of students in a curricular sequence of courses in English as a foreign language (EFL)

3 ___ Assessing average (net) gain in skills measured by the SLEP for students in academic EFL courses

4 ___ Assessing average (net) gain for students in intensive ESL or similar programs generally

5 ___ Screening applicants for admission to an academic program/institution

6 ___ Monitoring the progress of individual students

7 ___ Other (please describe briefly other uses, and elaborate briefly the use[s] indicated above):

_____
_____
_____
_____
_____
_____

Q2. Forms of the SLEP used (check all applicable):

1 ___ Form 1 2 ___ Form 2 3 ___ Form 3

Q3. _____ (No.). Please write in the approximate number of individuals tested during the past 12 months.

Q3a. Was the typical individual tested more than one time with the SLEP during the period?

___ 1 Yes 2 No

Q4. Estimated percentage of individuals tested in each of the following categories:

___ 1 Below 7th grade level (please answer Q4a)
___ 2 Grades 7 - 12
___ 3 College level or beyond (please answer Q4a)

Q4a. Please indicate the age/grade levels involved, and comment on the extent to which the SLEP appears to be suitable/unsuitable for the younger/older students, giving specific examples:

_____
_____
_____
_____
_____
_____

BEST COPY AVAILABLE

Q5a. Which of the following patterns of test use is most applicable for your setting?

1. The SLEP is the only standardized measure of English proficiency used (for the purposes indicated above)
2. The SLEP and locally developed tests or assessment procedures are being used
3. The SLEP, other standardized ESL/EFL tests, and locally developed tests or proficiency assessment procedures are used

Q5b. Please describe briefly typical testing practices indicating how the SLEP, other standardized tests, and/or locally developed tests are used to accomplish overall objectives.

Q5c. Please write in the names of up to three of the regularly or frequently used additional English proficiency tests.

Q5d. Have you been able to study the relationship between SLEP scores and direct measures of ESL/EFL students' ability to use English (e.g., English teacher's ratings of oral English proficiency, academic teachers ratings of students' ESL/EFL skills), and/or scores on other ESL/EFL tests?

1. Yes (please describe briefly below)
2. No (please comment briefly on your impressions regarding the foregoing, and reasons for them)

Q5d. The SLEP Total scaled score categories are specified below. Please check the score range that includes the average score typically obtained by students when initially tested.

1. - 32 2. 33-39 3. 40-46 4. 47-53 5. 54+

Q5e. Have you developed local norms for the SLEP (e.g., a table showing the percentage of students scoring at or below designated SLEP scores)?

1. Yes (If possible, please enclose a copy of your norms table and related description)
2. No

6. Approximately what percentage of all students tested are classifiable as (Total should be 100%)

1. ____ Permanent residents whose native language is not English
2. ____ Recent immigrants
3. ____ Refugees
4. ____ Undocumented
5. ____ ESL/EFL students in or planning to enter an English medium secondary or postsecondary level program
6. ____ Other (please describe)

Q7. Several general language groups are listed below. Please check all groups represented in your testing population. Please write in language groups that account for a higher proportion of the testing population than any of those listed.

1. ____ Spanish
2. ____ Japanese
3. ____ Chinese
4. ____ Arabic
5. ____ Korean
6. ____ Cambodian
7. ____ Vietnamese
8. ____ Thai
9. ____ Portugese
10. ____ Indonesian
11. ____ French
12. ____
13. ____
14. ____

Q8. What do you regard as the most positive features of the SLEP test (considering the uses indicated above)?

Q9. And what are its primary limitations, from the same perspective?

Q10. What do you regard as the most positive/negative features of the SLEP Test and Score Manual? What changes and/or additions to the Manual or the test itself would be most helpful to you?

See general question on the back of the cover memorandum. Thank you for our help.

70

71

BEST COPY AVAILABLE

Appendix C: Procedures Followed in Modifying Existing Addresses for the Current Survey

As indicated in the text, many of the addresses did not specify a pertinent "use-related" title or program. For example, many orders were placed by and shipped to school boards, school districts, institutional fiscal offices or agents, and so on. In order to provide a more specific target for the survey questionnaire, these general addresses were modified to include a plausible ESL-related recipient, as outlined below:

1. For U.S. and Canadian addresses involving district-level or board-level orders (e.g., Board of Education, School District No. 10, and so on), or orders placed through, or to be shipped to, a financial office (e.g., bursar, accounts payable), with no individual, departmental, or ESL program identification, a program (e.g., "ESL/Bilingual Program" for a district) or position/program (e.g., Director, ESL/Bilingual Program) was specified in the survey mailing.

2. In the case of individual schools, community colleges, four-year colleges and universities, or other institutions/ agencies for which no specific position/title/ESL program designation was available, a title/program designation (e.g., Director, ESL Proficiency Program) was added.

3. The "English Department" was targeted in the case of general addresses (other than Canadian) for orders from institutions clearly identified as schools, academies, colleges, and so on, outside the U.S. proper.

4. For orders placed through embassy, consular, or other governmental offices, a position title such as, "Adviser, ESL Proficiency Testing," or "Education Adviser," was specified.

The covering letter included a request that the questionnaire be forwarded to the appropriate individual or office.

Appendix D: Illustrative Items from the Sequential Tests of Educational Progress (STEP): Listening Comprehension and Reading Comprehension[**]

Brief descriptions of the listening comprehension and reading comprehension tests in the STEP series are provided in the following two pages. These tests are designed for use with native English-speaking students: the illustrative listening test material is for Grades 7-9; the reading test material is for Grades 4-6.

It is instructive to compare these items with those in the SLEP test (Appendix A). The SLEP items clearly are less cognitively demanding than are the STEP items. Accordingly, ESL students with average scores on SLEP can be expected to earn scores on a test like STEP that are below average relative to native-speaker norms.

See the *SLEP Test Manual* (e.g., ETS, 1987) for evidence indicating that **native English-speaking 7th graders can answer correctly almost all the SLEP items; see also Holloway (1984).**

It would be useful to determine the age/grade level at which SLEP items begin to represent a significant cognitive challenge for native English-speaking students.

** See ETS (1958).

The Listening Comprehension Tests

The members of the committee on the listening comprehension tests are:

Chairman—Althea Beery, Cincinnati Public Schools

Level 1 (Grades 13-14)
Seymour Eskow, Mohawk Valley Technical School, Utica, New York
*Ralph C. Leyden, Stephens College
Osmond E. Palmer, Michigan State University
Level 2 (Grades 10-12)
*John Caffrey, Los Angeles County Schools
Margaret Early, Syracuse University
Alice P. Sterner, Barringer High School, Newark, New Jersey
Level 3 (Grades 7-9)
William Grassl, University High School, University of Minnesota
*Stanley B. Kegler, University High School, University of Minnesota
Nathan A. Miller, Little River Junior High School, Miami, Florida
Level 4 (Grades 4-6)
*Ursula Hogan, Sacramento County Schools, Sacramento, California
Mildred Patterson, Public Schools, Wilmington, Delaware
Charlotte Wells, University of Missouri
*Members of planning committee, Althea Beery, Chairman

What is listening comprehension?

The student listens to be informed, to be inspired, to be convinced, or to be entertained. Whatever his purpose, it is important for the student to listen with understanding. Three goals, or levels of understanding, were established for intelligent listening:

1. **Comprehension** What is the plain sense of what is heard?

2. **Interpretation** What was the speaker trying to do? What were the implied meanings of the message? How does what is heard relate to other common knowledge?

3. **Evaluation** What are the weaknesses and strengths of the speaker's presentation? How valid is the message in the light of common knowledge?

A good listener is not a sponge, absorbing everything without discrimination. He listens critically and selectively. He remembers significant details, but not all details. More important, he remembers the speaker's main ideas and conclusions and appraises them critically. The development of such critical and selective listening is a goal with which schools are concerned. The STEP Listening Comprehension Tests are designed to measure the school's success in achieving this goal.

Criteria for selection of materials

How may children develop essential listening skills in school? At what grade levels should particular skills be emphasized? What sequence of development is desirable if children are to grow in listening ability?

In seeking answers to these questions as a basis for test development, these criteria were established:

- Listening situations should sample all types of listening familiar to students in their school experiences: directions and simple explanations, exposition, narration (both simple and figurative), argument and persuasion, aesthetic material.
- Language used should be real, that is, "language as it is heard," rather than language as it is read.
- Selections, and the questions based on them, must test a variety of skills and understandings, emphasizing selective memory and the ability to think about what is heard.

Skills tested

Basic listening skills were identified and organized around four aspects of what is communicated: main idea, significant details, organization of details, meaning of words. Since the skills are not isolated from one another, many test questions involve more than one skill. In most cases, however, it is possible to identify a basic skill required to answer the question.

I Plain-sense comprehension

1. To identify main ideas; to select a suitable title or to select a correct statement of the main idea or central theme.
2. To remember the significant details.
3. To remember the structure or simple sequence of ideas.
4. To demonstrate understanding of denotative meanings of important words.

II Interpretation (higher-level meanings)

1. To understand the implications of the main ideas; to understand what the speaker is trying to do; to see how the main ideas may reveal the speaker's attitudes and prejudices; to recognize the relationship of the speaker's statements to other ideas or to common knowledge.
2. To understand the implication of significant details; to understand how the details are pertinent to the speaker's purpose; to see how the details reveal the speaker's attitudes, biases, and prejudices; to see relationships among the details and their validity in the light of common knowledge.
3. To understand interrelationships among ideas and to understand the organizational pattern well enough to predict what is likely to follow.
4. To demonstrate understanding of connotative meanings of words; to infer meanings from the context; to understand how words are used to create a mood or an aesthetic feeling.

III Evaluation and application

1. To judge the validity and adequacy of the main idea; to distinguish fact from fancy; to distinguish probable fact from opinion and judgment.

BEST COPY AVAILABLE

Listening Comprehension

2 To judge the extent to which the supporting details accomplish their purpose; to distinguish among relevant and irrelevant details; to judge whether or not more information is needed to prove the speaker's point.

3 To evaluate the organization and development of what is said; to be aware of self-contradictions; to recognize the devices the speaker uses to influence the listener's thinking.

4 To judge whether or not the speaker has created an intended mood or effect—and if the speaker has failed, to understand why.

5 To recognize what the speaker wants the listener to do and to recognize ways in which the speaker's ideas may be applied properly in new situations.

Administration of the tests

The development of standardized tests in listening presents important problems of test administration. The alternative plans evaluated by the committee were to use tape recordings or to prescribe that the selections be read aloud by classroom teachers (or the test administrator). Each method has advantages and disadvantages; the evidence is not conclusive. The committee concluded that the evidence favored oral presentation by classroom teachers. This type of presentation makes the tests less expensive and does not require equipment for playing recordings.

Note: The term "auding" is acquiring wide currency to denote the entire process described here as listening comprehension. Sources of information about this relatively new field may be found in the *Review of Educational Research*, April 1955.

Sample of listening test material

Because listening comprehension is tested with fairly long dictated passages, there is space here for a sample from only one level. The examiner reads the passage once and then reads aloud both the questions and the possible answers. The student has before him a booklet that gives only the possible answers—and an answer sheet.

Level 3 (Grades 7-9) Reading time—1 min., 30 sec.

The examiner reads:

Here is the fourth selection. It is a speech by a student running for school office.

A students, B students, C students, D students, and my friends! As you know, I am running for the office of President of the Student Council. I'd like to tell you what I'll do if I'm elected. In the first place, I think several students ought to sit in on teachers' meetings. They settle too many things for us. I don't think that the teachers always know what's best for us.

In the second place, I'd like to see our Student Council do something. Take the business of the candy machine, for instance. Just because a couple of doctors and dentists don't like it doesn't mean we shouldn't have one. I think they are wrong. I think we should have one. Candy is good for us. It gives us energy, and I, for one, don't think it hurts either your teeth or your appetite. And if it does, so what? You save the lunch money and can go out on a date.

Last, you know that my opponents—and you'll hear from them in a minute—are two girls. Now, everybody says girls are smarter than boys. That might be true—but just because they're smarter doesn't mean they'll make better officers. In fact, I think girls are too smart and can't always get along with people because of that. Maybe we need somebody not so smart, but that can get along. That's me, fellow students—vote for me!

19 The speaker's principal objection to girls as school officers evidently is that they

A talk too much
B support the teacher's point-of-view
C are too smart to get along with people
D don't want a candy machine

20 It is likely that in the past the speaker has

E disagreed with the teachers' decisions
F disagreed with the opinions he has stated
G agreed with the doctor about the candy machine
H agreed with his other opponents about decisions of teachers

21 By saying "A students, B students, C students, D students, and my friends," the speaker is trying chiefly to get all the students

A at the top of the class to vote for him
B at the bottom of the class to vote for him
C in the school to vote for him
D who agree with him to vote for him

22 When the speaker used the word "opponents" he meant

E students from other schools
F students running against him
G the teachers
H doctors and dentists

23 Judging from his comments, how does the speaker feel about the opinions of experts?

A He pretends that the experts agree with him.
B He does not respect the experts if he disagrees with them.
C He pretends to treat the experts with respect.
D He follows expert advice unless he can prove that it is wrong.

The Reading Comprehension Tests

The members of the committee on reading comprehension are:

Chairman — Constance M. McCullough, San Francisco State College

Level 1 (Grades 13-14)
*Robert M. Bear, Dartmouth College
Philip Shaw, Brooklyn College
Macklin Thomas, Chicago City Junior College
Level 2 (Grades 10-12)
*Luella B. Cook, Minneapolis Public Schools
Dorothy E. McCullough, Tudor Hall School, Indianapolis
M. Myriam Page, Oakwood High School, Dayton, Ohio
Level 3 (Grades 7-9)
*Lauren L. Brink, University of Nevada
Helen F. Olson, Queen Anne High School, Seattle
Jerry L. Reed, Kepner Junior High School, Denver
Level 4 (Grades 4-6)
Harvey Alpert, University of Florida
Robert B. Simpson, San Francisco Public Schools
*George D. Spache, University of Florida

*Members of planning committee, Constance M. McCullough, Chairman

Purposes of reading

Of all the magic that education produces, there is little comparable to what happens when a child learns to read. And when the school helps the child to read with increasing skill and insight, it extends the magic by giving him a basic tool for understanding himself and his world. Whatever differences there may be regarding the goals of education —and there are many—the development of reading skills for all children remains the first of the three R's.

Possibly this is why more progress has been made in the measurement of reading comprehension than in other areas of educational testing. A number of educators and psychologists have made lists of abilities which are thought to add up to the ability to read well. The committee on reading had the advantage of acquaintance with the long history of analytical work in reading comprehension and knowledge of the types of testing materials that have been used successfully in the past.

The purpose of the STEP reading tests is to evaluate student ability to read new materials with comprehension, insight, and critical understanding. The task of the committee was not so much to explore new ground (as in the writing and listening tests) as it was to develop a plan which would take advantage of those current and past developments which are most closely related to the philosophy of the STEP program.

Criteria for selection of materials

Reading passages should be

- interesting to the pupils tested, and neither obviously dated nor offensive in any way.
- of a kind similar to those read by pupils in their ordinary school and life situations, but not likely to be familiar to students taking the tests.
- crucial in value.
- distributed in difficulty across several grades for tests at each level.
- more or less self-contained and representative of a variety of types of reading, a variety of fields and content, and a variety of media of written communication.

Types of materials

To test breadth of student development in reading skills, the selections represent a wide range of content, but the tests do not emphasize understanding of concepts or developed ability in any of the subject areas. Moreover, selections contain the information needed to answer the questions. However, the tests do measure the ease with which students read in the various content areas. (This difference in purpose can be seen by comparing reading test selections with items used in the science and social studies tests.)

What are the reading skills tested?

So far as possible, questions on each passage are distributed among five general categories of skills identified.

1. **Ability to understand direct statements made by the author:** to understand denotative meanings of words; to identify parallel statements; to recognize paraphrases; to recognize a correct statement of time sequence; to identify things mentioned most frequently.
2. **Ability to interpret and summarize the passage:** to select a suitable title; to identify the type of passage (fiction, history, etc.); to draw inferences from statements made by the author; to understand connotative meanings of words.
3. **Ability to see the motives of the author:** to be able to state the author's purpose; to understand why the author included or excluded certain things; to identify the tone of the passage.
4. **Ability to observe the organizational characteristics of the passage:** to recognize where divisions might come in a single long paragraph; to state the main topics of separate paragraphs; to understand the basis on which a passage is organized.
5. **Ability to criticize the passage with respect to its ideas, purposes, or presentation:** to judge if an argument is unsupported; to identify a valid objection not answered by the author; to judge effectiveness of devices used by the author (metaphor, example, etc.); to be aware of basic assumptions the author expects the reader to take for granted.

BEST COPY AVAILABLE

Reading Comprehension

Samples of reading test material

Reading passages were first classified according to grade level of difficulty. A set of passages representing the seven types of reading material was then selected for each level. Test items (questions) were designed to cover as many of the reading skills as possible. These items were then reviewed, revised as necessary, and arranged into whole tests.

The following samples from Level 4 (Grades 4-6) and Level 2 (Grades 10-12) illustrate some of the kinds of questions used. Both selections are classified as "opinion or interpretation" and require use of abilities in several categories.

Level 4 (Grades 4-6)

Dear Bill,

It was fun to be on the farm. Yesterday morning, Jack and I watched Aunt Mary make butter. She did not need to use all her cream to make butter. She sent most of the cream to the creamery.

I wish I were a farmer. I would take just a little cream for butter. Then I would use all the rest of the cream to make ice cream. Wouldn't that be fun?

I'm sorry you could not go to Jack's farm with me. I had the time of my life. Every day, Jack kept finding some new thing to do.

We rode Jack's horse. We worked around the barn. We fed the animals. We gave corn to the hogs in their pen. What a noise a hog can make. We gave hay to the horses and the sheep and the little lamb.

I came back to town yesterday. I must say good-by for now. Write soon.

Your cousin,
Betty

6 In this letter, Betty is trying to tell
F how to make butter
F what she did at the farm
G what horses eat
H how much noise a hog makes

* * *

8 Which of these things that Betty said tells best how she feels about living on a farm?
E We worked around the barn.
F I came back to town yesterday.
G I wish I were a farmer.
H We rode Jack's horse.

9 The letter is happy except where Betty is
A saying Bill couldn't come
B telling about riding the horse
C having to say good-by
D telling about the cream

10 Where does Betty live?
E In the mountains F On a farm
G Near the ocean H In a town

Level 2 (Grades 10-12)

In turn-of-the-century vaudeville, folding beds were favorite comedy props, but the many descendants of those early folding beds are no laughing matter. Today's smaller homes call for furniture that conserves space by serving more than one purpose, and the modern "convertibles" are going far toward satisfying that need. They can turn the most proper living room into a dormitory that will sleep nine people. Convertible furniture is giving American homemakers the imaginative engineering, improved design, and remarkable mass-production prices associated with home appliances. This development has provided the biggest home-furnishing news in recent years. In 1940, United States families spent about 22 million dollars for convertible sleep furniture; now, they are spending six times that amount for beds that hide in the living room during the day.

11 The information in this passage would be of interest to
A housewives B furniture manufacturers
C buyers for furniture stores D all of these

12 It is evident from the article that
E furniture designers are concentrating on the needs of small houses
F today's smaller houses require smaller furniture
G modern bedrooms will have to accommodate more than two persons
H old-fashioned furniture can be converted to fulfill today's requirements

13 Which of the following techniques does the author use to make his presentation of ideas effective?
A Supporting a statement with specific proof
B Giving figures
C Listing advantages D all of these

14 United States families buy convertible furniture today at an annual cost of
E over 100 million dollars
F 66 million dollars
G 22 million dollars
H 6 million dollars

15 In the sentence beginning in line 6 ("They can turn ..."), the author is
A adding an entirely new idea to his article
B illustrating the meaning of the preceding sentence
C generalizing from the preceding sentence
D making a general statement which will be followed by an example in the next sentence

11

BEST COPY AVAILABLE

Endnotes

1. The SLEP test was originally administered only in scheduled administrations at international test centers established by ETS, a practice that was discontinued in the early 1980s. However, since that time, SLEP has been continuously available for local administration and scoring, to qualified institutions, agencies, or individuals, for local administration, through a program that has come to be known as the SLEP School Services Program (ETS, 1991).

2. Two answer-sheet formats are offered: a three-ply format, in which sheets 2 and 3 record direct images of the correct responses only, and a single-sheet format that requires the use of scoring stencils.

3. Perspective regarding problems, issues, and practices in ESL proficiency assessment in secondary school settings was gained through discussions with individuals in the New Jersey Department of Education regarding the evaluation of federal/state-funded programs for students of Limited English Proficiency (LEP); also through a meeting with individuals responsible for ESL/bilingual programs in the Princeton (NJ) area, and the director of an ESL program for international students at a private secondary school in the same general area.

4. Individuals concerned with ESL/Bilingual programs in three New Jersey secondary school settings completed and commented on the draft, as did three university-based ESL program staff members. The draft was reviewed by the TOEFL program representative in Canada.

5. In the U.S. and Canada, ESL proficiency testing is mandated for use in evaluating certain ESL/bilingual programs sponsored by federal/state/provincial governments. This reason for test use was not anticipated for other countries, in which it was anticipated that the SLEP test might be used to assess gains in proficiency associated with academic "English as a foreign language" (EFL) instructional programs.

6. The principal source of records regarding "users" of the SLEP test was the general systems file maintained for fiscal accounting purposes. In many instances, orders for sets of test booklets (Form 1, Form 2, and/or Form 3) and related materials are placed by or through a business or purchasing office or agent. Neither the program in which the SLEP test is ultimately used nor specific name/title/program identification for the responsible test user is available. Moreover, systems files, intended primarily to meet <u>current</u> operational demands, are not designed to provide a consolidated, historical record of transactions by purchasers.

7. Traditionally, all orders for the SLEP test have been processed by the ETS (Princeton) office. In recent years, the TOEFL representative office in Canada has processed orders originating in that country.

8. The survey questionnaire did not ask for information regarding the type of answer sheet and scoring procedure employed. In relatively "high volume" contexts, scannable answer sheets and computerized scoring procedures undoubtedly were used (a practice that was specifically reported by only one respondent).

9. The SLEP School Services Program has made the test available " . . . for purchase and use . . . by postsecondary institutions, training agencies, educational consultants, and others engaged in legitimate testing activities" (e.g., ETS, 1988, p. 8). Use with ESL students whose age/grade placement is below the G7-12 range is indicated by informal feedback from the field.

10. As will be seen in a subsequent section, these items also were mentioned unfavorably by respondents in free responses identifying "positive" and "negative" features of SLEP, generally.

11. See Appendix B for differences in the detailed specification of purposes for testing--for "placement" and for "assessing average (net) gain"--in questionnaires for U.S.A/Canada and other locations, respectively, that are not directly pertinent here.

12. In subsequent correspondence with one respondent to the survey questionnaire, in a use-context in which placement decisions involved a composite of interview ratings, essay ratings and SLEP scores, it was learned that the variable of major interest for local "normative" purposes was the composite, not SLEP or other component elements--whose local distributions were well known.

13. Several respondents enclosed documents describing local studies and/or study outcomes; one respondent enclosed a report describing a study of change in test performance associated with intensive ESL instruction.

14. Strength of association was rarely characterized statistically; only two respondents reported a correlation coefficient to indicate strength of association between measures. In a number of instances, the "other variable(s)" involved were not explicitly described.

15. Illustrative responses in this category include the following verbatim comments by respondents who indicated that a study had been conducted: "Since SLEP is used for placement purposes, teachers' comments after placement are very important for any adjustments in placement" (from a U.S. university respondent). "Compare SLEP scores with academic grades and teacher judgment" (from a U.S. high school ESL teacher).

16. This is likely to be the case in most similar testing situations (see, for example, Cummins, 1983; the SLEP Test Manual [e.g., ETS, 1987, pp. 34-35]). Also, compare the listening comprehension and reading comprehension items in a test for G4-G9 native English speakers (Appendix D), with the corresponding SLEP items (Appendix A).

17. NCEs represent a transformation of percentile distributions to a standard scale (mean = 50, sd = 22) that permits "equal interval" comparisons regardless of score level. This index is widely used in conjunction with mandated models for assessing average (net) gain in test performance for students in federally or state funded remedial programs in the United States.

18. As put by one ESL/Bilingual supervisor: "Although SLEP is not a criterion referenced test, it would be helpful to know what underlying skills or curriculum goals, if any, are addressed by the test items." Expert classification of test items according to "skills/functions" appears to be feasible, and would permit useful extension of the information provided by SLEP.

19. None of the respondent's indicated precisely how SLEP was used to monitor the progress of individual students. However, one respondent expressed keen dissatisfaction with SLEP because some students had lower scores when posttested than they earned when pretested--a phenomenon that reflects factors subsumed under the rubric of "errors of measurement"--although the group as a whole apparently registered an average (net) gain. It would be useful to include in the SLEP Test Manual, a brief discussion of the problems associated with using a simple test-retest model for evaluating the progress of individual students (as opposed to the use of such a model for assessing average change).

20. Examples of available evidence bearing on SLEP's validity for use with college-level samples include the following:

(a) One survey respondent reported correlations averaging .63 between SLEP total score and professionally rendered ratings of oral English proficiency (based on formal

interviews) and ratings of writing samples, respectively, for Japanese students (N = 1,648) planning to enter a college-level English-medium program in Japan; other college-level SLEP users reported favorably on SLEP's usefulness (validity) for ESL assessment purposes.

(b) SLEP scores have been found to be relatively closely related to TOEFL scores in one sample of college-level students--correlations centering around .80 in a sample of students in a college-based intensive ESL program, with a mean of 519 on TOEFL and 55 (80th percentile) on SLEP (e.g., ETS, 1988).

(c) A study (Butler, 1989) of SLEP performance of ESL students in member institutions of the Los Angeles Community College District (LACCD) indicated that SLEP performance of independently established proficiency-level groups varied systematically with placement level, and that the test items were at an appropriate level of difficulty for the students involved.

(d) A study (Rudmann, 1991) involving ESL students at Irvine Valley (CA) Community College, found that SLEP scores were related positively (average levels approximately .40) to grades earned in English courses; this despite the fact that the students were assigned to the respective courses on the basis of SLEP scores, with attendant restriction of range on the test within the respective course-level samples.

Findings such as the foregoing, constitute what appears to be "conceptually persuasive" evidence that SLEP can be expected to provide reliable and valid discrimination in samples of college-level ESL students. Further evidence is needed--validity assessment can never be considered "complete."

21. ESL proficiency assessment, mandated in connection with governmentally funded programs for students with limited English proficiency, typically must be conducted using only "approved" tests and procedures. For example, New Jersey administrative codes specify that "an English language proficiency test, in the areas of listening, speaking, reading, and writing, must be administered to those pupils with another language in their background . . .. The Language Assessment Battery (LAB, 1982) and the Maculaitis Assessment Program (MAC, 1982) are the tests used for this purpose. However, other language proficiency tests may be used, as long as the tests have validity and reliability, measure the areas of listening, speaking, reading, and writing, and have been aligned to the state norms established for the LAB and MAC tests" (New Jersey State Department of Education, 1990: p.

10). Remedial work may be focused primarily on particular skills (e.g., reading comprehension). "Norm-referenced models" have been developed for use in evaluating programs. ESL proficiency testing (along with assessment of age/grade appropriate subject-matter attainment in English and/or L1) is conducted locally, using tests selected by local districts from lists of state-approved tests. **Currently approved tests may lose "approved" status if their norms are more than ten years old (see, for example, New Jersey State Department of Education, 1990, p. 130).**

22. Respondents in a number of SLEP use contexts indicated that they would like more information regarding the strength and consistency of relationships between SLEP scores and scores on TOEFL. The TOEFL appears to be the "ultimate" challenge for ESL students in many SLEP use contexts--indeed one measurable goal of instruction mentioned by several respondents was the attainment of a particular TOEFL score. For the informed guidance of practitioners, information is needed regarding the typical level and range of TOEFL performance that can be expected (concurrently or after some designated period of instruction) for examinees with particular scores on SLEP. Accordingly, it would be useful to collect data needed to extend evidence regarding SLEP/TOEFL relationships **in samples from both secondary-level and college-level SLEP use contexts.** In any event, in reporting on relationships to test users--e.g., indicating expected TOEFL scores for given SLEP score ranges, as in the SLEP Test Manual (e.g., ETS, 1991, Table 16, p. 27)--an "expectancy table" format, rather than a simple "table of equivalents," should be used. One survey respondent expressed considerable dismay upon discovering that the actual TOEFL scores earned by her students were frequently considerably at variance with the "equivalents" indicated in the *SLEP Test Manual*. Seeing the scatter of TOEFL scores for examinees in designated SLEP score ranges should help users to form realistic expectations.

23. ESL professionals in college-level settings, interviewed a decade ago (Hale and Hinofotis, 1981), also reportedly stressed " . . . the need to assess productive as well as receptive skills" (p. 9) for placement purposes.

24. The results of criterion-related validity studies involving "common criteria" can be expected to provide useful general guidelines for test interpretation, based on the results of studies that have been designed explicitly to link level of performance on indirect, norm-referenced measures to quasi-absolute proficiency scales, using ratings of classroom ESL teachers (e.g., from the TOEFL testing context, see Boldt, Larsen-Freeman, Camp, & Levin, in press; and from the TOEIC testing context, see Wilson, 1991).

25. The reference here is not to the development of a test with multiple scores to be employed in identifying strengths and weakness of individuals, but to the use of average scores on items designed to measure particular skills to identify skill-areas requiring more/less emphasis in instruction, or to compare groups with respect to profiles of skills--that is, information that can be useful for evaluating or planning instruction. As noted by Hale and Hinofotis (1981: p 20): "It is possible to employ a basically integrative approach with tests focusing on the assessment of the major skills in an appropriate context and, at the same time, to provide a breakdown by subskills or objectives within those major skills."

26. This approach was employed by Clark and Swinton (1979) in their study concerned with the development of the Test of Spoken English (TSE). The final selection of TSE items was based in part on patterns of correlation with ratings of oral English proficiency.

27. See Hale and Hinofotis (1981--pp. 20-22) for illustrative analytic approaches to the problem of providing ". . . a breakdown by subskills or objectives within . . . major skill areas (tapped by an integrative test)."

References

Board of Education, City of New York (1982). Language assessment battery. NY: Author.

Boldt, R. F., Larsen-Freeman, D., Reed, M., and Courtney, R. (1992). Descriptors of TOEFL score levels (TOEFL Research Report No. 41). Princeton, NJ: Educational Testing Service.

Butler, F. H. (1989). Placement testing for a district-wide ESL matriculation model: Los Angeles Community College District (Project Report). Los Angeles, CA: Los Angeles Community College District.

Clark, J. L. D., and Swinton, S. S. (1979). An exploration of speaking proficiency measures in the TOEFL context (TOEFL Research Report No. 4). Princeton, NJ: Educational Testing Service.

Cummins, J. (1983). Conceptual and linguistic foundations of language assessment. In S. S. Seidnor (Ed.), Issues of Language Assessment, Vol. II: Language Assessment and Curriculum Planning (7-16). Rosslyn, VA: National Clearinghouse on Bilingual Education.

Educational Testing Service (1991). SLEP Test Manual. Princeton, NJ: Author.

Educational Testing Service (1988). SLEP Test Manual. Princeton, NJ: Author.

Educational Testing Service (1987). SLEP Test Manual. Princeton, NJ: Author

Educational Testing Service (1958). A brief: Sequential tests of educational progress. Princeton, NJ: Author.

Hale, G. J, and Hinofotis, F. B. (1981). New directions in English language testing (draft report). Princeton, NJ: Educational Testing Service.

Holloway, D. M. (1984). Validation study of SLEP test. Unpublished master's thesis, University of Florida, Gainesville, FL.

Maculaitis, J. D. (1982). Technical manual for Maculaitis assessment program: A multi-purpose test for non-native speakers of English in kindergarten through twelfth grade (MAC:K-12). San Francisco: The Alemany Press.

New Jersey State Department of Education (1990). Guidelines for development of program plan and evaluation summary: Bilingual/ESL programs and English language services, fiscal year 1991. Trenton, NJ: Author.

New Jersey State Department of Education (1989). Norming study of the language assessment battery and the Maculaitis assessment program: Practitioner's report. Trenton, NJ: Author.

Rudmann, J. (1991). A study of the secondary level English proficiency test (SLEP) at Irvine Valley College. Irvine, CA: Author.

Stansfield, C. (1984). Reliability and validity of the secondary level English proficiency test, System, 12, 1-12.

Tillberg, R. (Oct. 22, 1991). Personal communication.

Wilson, K. M. (1991). Validity of the test of English for proficiency advancement (TOEPA) for predicting teacher's ratings in diverse settings (Draft project report). Princeton, NJ: Educational Testing Service.

BEST COPY AVAILABLE